

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2000, 01:33:37 ; Search time 9.32 Seconds
(without alignments)
10.790 Million cell updates/sec

Title: US-08-894-356C-21

Perfect score: 40

Sequence: 1 DFGWCK 6

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA:*
- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/1aa/6_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	448	1	US-08-207-904-2
2	40	100.0	448	1	Sequence 2, Appl
3	31	77.5	45	2	Sequence 17, Appl
4	31	77.5	45	2	Sequence 51, Appl
5	30	75.0	33	2	Sequence 6, Appl
6	30	75.0	33	2	Sequence 49, Appl
7	30	75.0	33	2	Sequence 54, Appl
8	30	75.0	259	1	Sequence 10, Appl
9	30	75.0	259	1	Sequence 10, Appl
10	30	75.0	259	1	Sequence 10, Appl
11	30	75.0	259	1	Sequence 10, Appl
12	30	75.0	259	2	Sequence 10, Appl
13	30	75.0	259	4	Sequence 10, Appl
14	30	75.0	260	1	Sequence 10, Appl
15	30	75.0	260	1	Sequence 10, Appl
16	30	75.0	292	1	Sequence 81, Appl
17	30	75.0	292	1	Sequence 81, Appl
18	30	75.0	357	1	Sequence 2, Appl
19	30	75.0	357	1	Sequence 9, Appl
20	30	75.0	357	2	Sequence 4, Appl
21	30	75.0	370	2	Sequence 2, Appl
22	30	75.0	395	1	Sequence 3, Appl
23	30	75.0	395	2	Sequence 3, Appl
24	30	75.0	463	3	Sequence 1, Appl
25	29	72.5	13	1	Sequence 8, Appl
26	29	72.5	13	2	Sequence 8, Appl
27	29	72.5	14	2	Sequence 24, Appl
28	29	72.5	23	1	Sequence 132, App

29 29 72.5 23 1 US-08-471-052A-132 Sequence 132, App
30 29 72.5 23 1 US-08-189-331-132 Sequence 132, App
31 29 72.5 23 2 US-08-471-939-132 Sequence 132, App
32 29 72.5 23 2 US-08-471-800-132 Sequence 132, App
33 29 72.5 23 2 US-08-471-068-132 Sequence 132, App
34 29 72.5 60 3 US-08-468-011A-12 Sequence 12, Appl
35 29 72.5 154 2 US-08-449-644-7 Sequence 7, Appl
36 29 72.5 154 2 US-08-087-244A-7 Sequence 7, Appl
37 29 72.5 208 3 US-08-906-769-151 Sequence 151, App
38 29 72.5 208 3 US-08-906-616-151 Sequence 151, App
39 29 72.5 235 1 US-08-591-989-2 Sequence 2, Appl
40 29 72.5 242 2 US-08-512-955-4 Sequence 4, Appl
41 29 72.5 247 2 US-08-997-080-135 Sequence 135, App
42 29 72.5 247 2 US-08-997-362-135 Sequence 135, App
43 29 72.5 299 2 US-08-997-080-124 Sequence 124, App
44 29 72.5 299 2 US-08-997-362-124 Sequence 124, App
45 29 72.5 323 1 US-08-591-989-5 Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-207-904-2
; Sequence 2, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie
; APPLICANT: Crossland, Lyle D.
; TITLE OF INVENTION: Anther-Specific cDNA Sequences, Genomic
; TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: CGC 1624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 448 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-207-904-2

Query Match 100.3%; Score 40; DB 1; Length 448;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DFGWCK 6
|||||

Db 392 DFGWGK 397

RESULT 2

US-08-904-17
; Sequence 17, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie
; TITLE OF INVENTION: Antisense Specific cDNA Sequences, Genomic
; TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: CGC 1624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 448 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-207-904-17

Query Match 100.0%; Score 40; DB 1; Length 448;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DFGWGK 6

Db 392 DFGWGK 397

RESULT 3

US-08-726-306A-51
; Sequence 51, Application US/08726306A
; Patent No. 5958684
; GENERAL INFORMATION:
; APPLICANT: van Leeuwen, Frederik Willem
; APPLICANT: Burbach, Johannes Peter Henri
; APPLICANT: Grosveld, Franklin G.
; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
; NUMBER OF SEQUENCES: 189
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1 Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US

ZIP: 02111
COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,306A
; FILING DATE: 02-Oct-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 95/20080.4
; FILING DATE: 02-Oct-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,832
; FILING DATE: 01-Jan-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-726-306A-51

Query Match 77.5%; Score 31; DB 2; Length 45;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DFGWGK 6

Db 1 DKGWGK 6

RESULT 4

US-08-936-135-6
; Sequence 6, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341

TELEFAX: (650) 343-4342
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 923 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-936-135-6

Query Match 77.5%; Score 31; DB 3; Length 923;
 Best Local Similarity 80.0%; Pred. No. 6.1e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFGWG 5
 Db 651 EFGWG 655

RESULT 5
 US-08-031-538-49
 ; Sequence 49, Application US/08031538
 ; Patent No. 5968817
 ; GENERAL INFORMATION:
 ; APPLICANT: Sutcliffe, J Gregor
 ; APPLICANT: Erlander, Mark G
 ; APPLICANT: Lovenberg, Timothy W
 ; TITLE OF INVENTION: HUMAN SEROTONIN RECEPTORS, DNA ENCODING
 ; NUMBER OF SEQUENCES: 73
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: The Scripps Research Institute, Office of
 ; ADDRESSEE: Patent Counsel
 ; STREET: 10666 No. 5968817th Torrey Pines Road, TPC 8
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/031,538
 FILING DATE: 19930315
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Fitting, Thomas
 REGISTRATION NUMBER: 34,163
 REFERENCE/DOCKET NUMBER: TSR5099P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-554-2937
 TELEFAX: 619-554-6312
 INFORMATION FOR SEQ ID NO: 49:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 33 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-031-538-49

Query Match 75.0%; Score 30; DB 2; Length 33;
 Best Local Similarity 80.0%; Pred. No. 4.1;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 FGWKG 6
 Db 5 FGWGE 9

RESULT 7
 US-07-901-707-10
 ; Sequence 10, Application US/07901707
 ; Patent No. 5376546
 ; GENERAL INFORMATION:
 ; APPLICANT: Bernhardt, Susan L.
 ; APPLICANT: Better, Marc D.
 ; APPLICANT: Carroll, Steve F.
 ; APPLICANT: Lane, Julie A.
 ; TITLE OF INVENTION: Materials Comprising and Methods of
 ; NUMBER OF SEQUENCES: 57
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
 ; ADDRESSEE: Bicknell
 ; STREET: Two First National Plaza, 20 South Clark
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA

Qy 2 FGWKG 6
 Db 5 FGWGE 9

RESULT 6
 US-08-031-538-54
 ; Sequence 54, Application US/08031538
 ; Patent No. 5968817
 ; GENERAL INFORMATION:
 ; APPLICANT: Sutcliffe, J Gregor
 ; APPLICANT: Erlander, Mark G
 ; APPLICANT: Lovenberg, Timothy W
 ; TITLE OF INVENTION: HUMAN SEROTONIN RECEPTORS, DNA ENCODING
 ; NUMBER OF SEQUENCES: 73
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: The Scripps Research Institute, Office of
 ; ADDRESSEE: Patent Counsel
 ; STREET: 10666 No. 5968817th Torrey Pines Road, TPC 8
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/031,538
 FILING DATE: 19930315
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Fitting, Thomas
 REGISTRATION NUMBER: 34,163
 REFERENCE/DOCKET NUMBER: TSR5099P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-554-2937
 TELEFAX: 619-554-6312
 INFORMATION FOR SEQ ID NO: 54:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 33 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-031-538-54

Query Match 75.0%; Score 30; DB 2; Length 33;
 Best Local Similarity 80.0%; Pred. No. 4.1;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 FGWKG 6
 Db 5 FGWGE 9

RESULT 7
 US-07-901-707-10
 ; Sequence 10, Application US/07901707
 ; Patent No. 5376546
 ; GENERAL INFORMATION:
 ; APPLICANT: Bernhardt, Susan L.
 ; APPLICANT: Better, Marc D.
 ; APPLICANT: Carroll, Steve F.
 ; APPLICANT: Lane, Julie A.
 ; TITLE OF INVENTION: Materials Comprising and Methods of
 ; NUMBER OF SEQUENCES: 57
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
 ; ADDRESSEE: Bicknell
 ; STREET: Two First National Plaza, 20 South Clark
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA

ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/901,707
FILING DATE: 19920619
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: No. 5376546and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27129/30910
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-5750
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 259 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-901-707-10

Query Match 75.0%; Score 30; DB 1; Length 259;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DFGWCK 6
Db 229 DFGFGK 234

RESULT 8
US-07-988-430-10
Sequence 10, Application US/07988430
Patent No. 5416202
GENERAL INFORMATION:
APPLICANT: Bernhard, Susan L.
APPLICANT: Better, Marc D.
APPLICANT: Carroll, Stephen F.
APPLICANT: Lane, Julie A.
APPLICANT: Lei, Shau-Ping
TITLE OF INVENTION: Materials Comprising and Methods of
Preparation and Use for Ribosome-Inactivating Proteins
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
Bicknell
STREET: Two First National Plaza, 20 South Clark
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/988,430
FILING DATE: 19921209
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,707
FILING DATE: 19-JUN-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: No. 5416202and, Greta E.
REGISTRATION NUMBER: 35302
REFERENCE/DOCKET NUMBER: 31133
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 259 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-988-430-10

Query Match 75.0%; Score 30; DB 1; Length 259;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DFGWCK 6
Db 229 DFGFGK 234

RESULT 9
US-08-425-336-10
Sequence 10, Application US/08425336
Patent No. 5621083
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
APPLICANT: Carroll, Stephen F.
APPLICANT: Studnika, Gary M.
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
Proteins
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,336
FILING DATE: 18-APR-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/064,691
FILING DATE: 12-MAY-1993
APPLICATION NUMBER: US 07/901,707
FILING DATE: 19-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Thomas C.
REGISTRATION NUMBER: P-36,989
REFERENCE/DOCKET NUMBER: 31394
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 259 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-425-336-10

Query Match 75.0%; Score 30; DB 1; Length 259;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DFGWGK 6
Db 229 DFGFGK 234

RESULT 10
US-08-488-113B-10
; Sequence 10, Application US/08488113B
; Patent No. 5744580
; GENERAL INFORMATION:
; APPLICANT: Better, Marc D.
; APPLICANT: Carroll, Stephen F.
; APPLICANT: Studnika, Gary M.
; TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held & Malloy, Ltd.
; STREET: 500 West Madison Street, 34th floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,113B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/425,336
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,691
; FILING DATE: 12-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,430
; FILING DATE: 09-DEC-1992
; APPLICATION NUMBER: US 07/901,707
; FILING DATE: 19-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,567
; FILING DATE: 04-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A
; TELEPHONE: 312/707-8889
; TELEX: 650 388-1248
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 259 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-488-113B-10

Query Match 75.0%; Score 30; DB 1; Length 259;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DFGWGK 6
Db 229 DFGFGK 234

RESULT 11
US-08-477-484B-10
; Sequence 10, Application US/08477484B
; Patent No. 5756699
; GENERAL INFORMATION:
; APPLICANT: Better, Marc D.
; APPLICANT: Carroll, Stephen F.
; APPLICANT: Studnika, Gary M.
; TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held & Malloy, Ltd.
; STREET: 500 West Madison Street, 34th floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,484B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/425,336
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,691
; FILING DATE: 12-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,430
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,707
; FILING DATE: 19-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,567
; FILING DATE: 04-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A
; TELEPHONE: 312/707-8889
; TELEX: 650 388-1248
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 259 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-484B-10

Query Match 75.0%; Score 30; DB 1; Length 259;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFGWCK 6

Db 229 DFGGCK 234

RESULT 12

US-08-646-360-10

; Sequence 10, Application US/08646360

; Patent No. 5837491

; GENERAL INFORMATION:

; APPLICANT: Better, Marc D.

; APPLICANT: Carroll, Stephen F.

; APPLICANT: Studnika, Gary M.

; TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating

; TITLE OF INVENTION: Proteins

; NUMBER OF SEQUENCES: 173

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McAndrews, Held & Malloy, Ltd.

; STREET: 500 West Madison Street, 34th floor

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60661

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/646,360

; FILING DATE: 13-MAY-1996

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/05348

; FILING DATE: 12-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/064,691

; FILING DATE: 12-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/988,430

; FILING DATE: 09-DEC-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: McNicholas, Janet M.

; REGISTRATION NUMBER: 32,918

; REFERENCE/DOCKET NUMBER: 200-70.P4

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312/707-8889

; TELEFAX: 312/707-9155

; TELEX: 650 388-1248

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 259 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-646-360-10

Query Match

Best Local Similarity 75.0%; Score 30; DB 2; Length 259;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFGWCK 6

Db 229 DFGGCK 234

RESULT 13

PCT-US92-09487-10

; Sequence 10, Application PC/TUS9209487

; GENERAL INFORMATION:

; APPLICANT: Bernhard, Susan L.

; APPLICANT: Better, Marc D.

; APPLICANT: Carroll, Stephen F.

; APPLICANT: Lane, Julie A.

; APPLICANT: Lei, Shau-Ping

; TITLE OF INVENTION: Materials Comprising and Methods of

; TITLE OF INVENTION: Preparation and Use for Ribosome-Inactivating Proteins

; NUMBER OF SEQUENCES: 101

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

; ADDRESSEE: Bicknell

; STREET: Two First National Plaza, 20 South Clark

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/09487

; FILING DATE: 19921104

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/901,707

; FILING DATE: 19-JUN-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/787,567

; FILING DATE: 04-NOV-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Noland, Greta E.

; REGISTRATION NUMBER: 35302

; REFERENCE/DOCKET NUMBER: 31133

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (312) 346-5750

; TELEFAX: (312) 984-9740

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 259 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; PCT-US92-09487-10

Query Match

Best Local Similarity 75.0%; Score 30; DB 4; Length 259;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFGWCK 6

Db 229 DFGGCK 234

RESULT 14

US-08-378-761A-72

; Sequence 72, Application US/08378761A

; Patent No. 5635384

; GENERAL INFORMATION:

; APPLICANT: WALSH, TERENCE A

; APPLICANT: HEY, TIMOTHY D

; APPLICANT: MORGAN, ALICE ER

; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE

;; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
;; TITLE OF INVENTION: USING
;; NUMBER OF SEQUENCES: 81
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ANDREA T. BORUCKI
;; STREET: 9330 ZIONSVILLE ROAD
;; CITY: INDIANAPOLIS
;; STATE: IN
;; COUNTRY: US
;; ZIP: 46268
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/378,761A
;; FILING DATE: 26-JAN-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BORUCKI, ANDREA T
;; REGISTRATION NUMBER: 33651
;; REFERENCE/DOCKET NUMBER: 38272B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (317) 337-4846
;; INFORMATION FOR SEQ ID NO: 72:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 260 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-378-761A-72

Query Match 75.0%; Score 30; DB 1; Length 260;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFGWGK 6
|||:|
Db 229 DFGFGK 234

RESULT 15
US-08-485-286-72
;; Sequence 72, Application US/08485286
;; Patent No. 5646026
;; Patent No. 5646026 5646119
;; GENERAL INFORMATION:
;; APPLICANT: WALSH, TERENCE A
;; APPLICANT: HEY, TIMOTHY D
;; APPLICANT: MORGAN, ALICE ER
;; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
;; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
;; NUMBER OF SEQUENCES: 81
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ANDREA T. BORUCKI
;; STREET: 9330 ZIONSVILLE ROAD
;; CITY: INDIANAPOLIS
;; STATE: IN
;; COUNTRY: US
;; ZIP: 46268
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/485,286
;; FILING DATE:
;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/378761
;; FILING DATE: 26-JAN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BORUCKI, ANDREA T
;; REGISTRATION NUMBER: 33651
;; REFERENCE/DOCKET NUMBER: 38272B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (317) 337-4846
;; INFORMATION FOR SEQ ID NO: 72:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 260 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-485-286-72

Query Match 75.0%; Score 30; DB 1; Length 260;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFGWGK 6
|||:|
Db 229 DFGFGK 234

Search completed: October 20, 2000, 01:35:57
Job time: 140 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2000, 01:34:22 ; Search time 11.52 Seconds
(without alignments)
33.052 Million cell updates/sec

Title: US-08-894-356c-21

Perfect score: 40

Sequence: 1 DFGWGK 6

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR.65:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	443	2	T45574
2	40	100.0	450	2	T45573
3	40	100.0	451	2	T00527
4	40	100.0	455	2	T09666
5	40	100.0	461	2	T03274
6	40	100.0	461	2	T00918
7	40	100.0	475	2	T45961
8	35	90.0	430	2	T46216
9	35	87.5	442	2	T10718
10	35	87.5	445	2	T10717
11	35	87.5	445	2	T10719
12	35	87.5	446	2	T10711
13	35	87.5	485	2	T33865
14	35	87.5	572	2	T01056
15	34	85.0	113	2	S30515
16	34	85.0	129	2	S60978
17	34	85.0	216	2	T02913
18	34	85.0	519	2	S54300
19	34	85.0	667	2	G69723
20	34	85.0	754	2	T47886
21	33	82.5	730	2	A75486
22	32	80.0	234	2	T36369
23	32	80.0	299	2	T36787
24	32	80.0	346	2	T10173
25	32	80.0	367	2	T06780
26	32	80.0	409	2	T19688
27	32	80.0	428	2	T48008
28	31	77.5	106	2	B64350
29	31	77.5	110	1	F71129

30	31	77.5	110	2	F75034	hypothetical prote
31	31	77.5	112	2	TJ0151	biphenyl-2,3-diol
32	31	77.5	158	2	S35201	serine proteinase
33	31	77.5	215	2	G70342	hypothetical prote
34	31	77.5	261	2	S74761	hypothetical prote
35	31	77.5	291	1	B53419	biphenyl-2,3-diol
36	31	77.5	293	1	DAPSFC	biphenyl-2,3-diol
37	31	77.5	296	2	JW0103	azarene carbazole
38	31	77.5	299	2	A57264	biphenyl-2,3-diol
39	31	77.5	299	2	A28718	2,3-dihydroxybiphe
40	31	77.5	299	2	T31282	biphenyl-2,3-diol
41	31	77.5	316	2	F69978	sugar-phosphate de
42	31	77.5	335	2	T48319	hypothetical prote
43	31	77.5	371	2	B65502	alcohol dehydrogen
44	31	77.5	393	2	S02185	uroporphyrin-III C
45	31	77.5	412	2	T40155	mannose-6-phosphat

ALIGNMENTS

RESULT 1
T45574
anthranilate N-hydroxycinnamoyl/benzoyltransferase-like protein - Arabidopsis thalian
N:Alternate names: protein F1IC1.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
R:Barques, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.;
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223007
A:Accession: T45574
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-443 <BAR>
A:Cross-references: EMBL:AL132976
A:Experimental source: cultivar Columbia; BAC clone F1IC1
C:Genetics:
A:Map position: 3
A:Note: F1IC1.120

Query Match
Best Local Similarity 100.0%; Score 40; DB 2; Length 443;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 DFGWGK 6
Db 386 DFGWGK 391

RESULT 2
T45573
anthranilate N-hydroxycinnamoyl/benzoyltransferase-like protein - Arabidopsis thalian
N:Alternate names: protein F1IC1.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
R:Barques, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.;
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223007
A:Accession: T45573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-450 <BAR>
A:Cross-references: EMBL:AL132976
A:Experimental source: cultivar Columbia; BAC clone F1IC1
C:Genetics:
A:Map position: 3
A:Note: F1IC1.110

Query Match 100.0%; Score 40; DB 2; Length 450;

Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
|||||
DB 391 DFGWGK 396

RESULT 3

T00527
hypothetical protein T20K24.8 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 23-Apr-1999
C:Accession: T00527
R:Rounsley, S.D.; Kaul, S.; Jin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, R. submitted to the EMBL Data Library, July 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A:Reference number: Z14167
A:Accession: T00527
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-451 <ROU>
A:Cross-references: EMBL:AC002392; NID:g3176701; PID:g3176709
C:Genetics:
A:Map position: 2
A:introns: 322/1
A:Note: T20K24.8

Query Match
Best Local Similarity 100.0%; Score 40; DB 2; Length 451;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
|||||
DB 397 DFGWGK 402

RESULT 4

T09666
probable anthranilate N-benzoyltransferase (EC 2.3.1.144) - muskmelon (fragment)
N:Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase
C:Species: Cucumis melo (muskmelon)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
R:Aggelis, A.; John, I.; Karvouni, Z.; Grierson, D.
Plant Mol. Biol. 33, 313-322, 1997
A:Title: Characterization of two cDNA clones for mRNAs expressed during ripening of melo
A:Reference number: Z16810; MUID:97188564
A:Accession: T09666
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-455 <AGG>
A:Cross-references: EMBL:T09521; NID:e991794; PID:e224032
A:Experimental source: cultivar Cantaloupe charientais; pericarp of ripe fruit
C:Keywords: acyltransferase; coenzyme A

Query Match
Best Local Similarity 100.0%; Score 40; DB 2; Length 455;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
|||||
DB 374 DFGWGK 379

RESULT 5

T03274
har201 protein, hypersensitivity-related - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 07-May-1999

C:Accession: T03274
R:Czernik, P.; Huang, H.C.; Marco, Y.
Plant Mol. Biol. 31, 255-265, 1996

A:Title: Characterization of har201 and har515, two tobacco genes preferentially expr
A:Reference number: Z14876; MUID:96343929

A:Accession: T03274
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-460 <CZE>
A:Cross-references: EMBL:X95343; NID:g1171576; PID:e220213
A:Experimental source: cultivar bottom special
C:Genetics:
A:Gene: har201

Query Match
Best Local Similarity 100.0%; Score 40; DB 2; Length 460;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
|||||
DB 382 DFGWGK 387

RESULT 6

T00918
hypothetical protein F21B7.32 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T00918
R:Shin, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.;
eologis, A.; Ecker, J.R.
submitted to the EMBL Data Library, January 1998
A:Description: Genomic sequence for Arabidopsis thaliana BAC F21B7.
A:Reference number: Z14208
A:Accession: T00918
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-461 <SHI>
A:Cross-references: EMBL:AC002560; NID:g2618677; PID:g2809263; GSPDB:GN00059; ATSP:F2
C:Genetics:
A:Gene: ATSP:F21B7.32
A:Map position: 1

Query Match
Best Local Similarity 100.0%; Score 40; DB 2; Length 461;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
|||||
DB 400 DFGWGK 405

RESULT 7

T45961
anthranilate N-benzoyltransferase-like protein - Arabidopsis thaliana
N:Alternate names: protein F7J8.190
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
R:Bevan, M.; Zimmermann, W.; Gruenelsen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.;
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23018
A:Accession: T45961
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-475 <BEV>
A:Cross-references: EMBL:AL137189
A:Experimental source: cultivar Columbia; BAC clone F7J8
C:Genetics:
A:Map position: 5
A:Note: F7J8.190

Query Match 100.0%; Score 40; DB 2; Length 475;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DFGWG 6
Db 411 DFGWG 416

RESULT 8
T46216
hypothetical protein T8P19.230 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 17-Mar-2000
C:Accession: T46216
R:Chouane, N.; Robert, C.; Brotlier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223008
A:Accession: T46216
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-430 <CHO>
A:Cross-references: EMBL:AL133315
A:Experimental source: cultivar Columbia; BAC clone T8P19
C:Genetics:
A:Map position: 3
A:Introns: 144/3
A:Note: T8P19.230
C:Superfamily: Arabidopsis CER2 protein

Query Match 90.0%; Score 36; DB 2; Length 430;
Best Local Similarity 83.3%; Pred. No. 42;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DFGWG 6
Db 378 DFGWG 383

RESULT 9
T10718
anthranilate N-benzoyltransferase (EC 2.3.1.144) (clone pchbcl1) - clove pink (fragment
N:Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase
C:Species: Dianthus caryophyllus (clove pink)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C:Accession: T10718
R:Yang, Q.; Reinhard, K.; Schiltz, E.; Matern, U.
Plant Mol. Biol. 35, 777-789, 1997
A:Title: Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:
lus L.
A:Reference number: 217095; MUID:98088004
A:Accession: T10718
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-442 <YAN>
A:Cross-references: EMBL:Z84384; NID:e1019563; PID:e294140
C:Function:
A:Description: catalyzes the synthesis of anthranilate
A:Pathway: phycoalexin biosynthesis
C:Keywords: acyltransferase; coenzyme A

Query Match 87.5%; Score 35; DB 2; Length 442;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DFGWG 5
Db 389 DFGWG 393

RESULT 10
T10717
anthranilate N-benzoyltransferase (EC 2.3.1.144) (clone pchbcl1) - clove pink
N:Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase
C:Species: Dianthus caryophyllus (clove pink)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C:Accession: T10717
R:Yang, Q.; Reinhard, K.; Schiltz, E.; Matern, U.
Plant Mol. Biol. 35, 777-789, 1997
A:Title: Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-Co
lus L.
A:Reference number: 217095; MUID:98088004
A:Accession: T10717
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-445 <YAN>
A:Cross-references: EMBL:Z84383; NID:e1019562; PID:e294139
C:Function:
A:Description: catalyzes the synthesis of anthranilate
A:Pathway: phycoalexin biosynthesis
C:Keywords: acyltransferase; coenzyme A

Query Match 87.5%; Score 35; DB 2; Length 445;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DFGWG 5
Db 392 DFGWG 396

RESULT 11
T10719
anthranilate N-benzoyltransferase (EC 2.3.1.144) (clone pchbcl3) - clove pink
N:Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase
C:Species: Dianthus caryophyllus (clove pink)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C:Accession: T10719
R:Yang, Q.; Reinhard, K.; Schiltz, E.; Matern, U.
Plant Mol. Biol. 35, 777-789, 1997
A:Title: Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-Co
lus L.
A:Reference number: 217095; MUID:98088004
A:Accession: T10719
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-445 <YAN>
A:Cross-references: EMBL:Z84571; NID:e1019567; PID:e294143
C:Function:
A:Description: catalyzes the synthesis of anthranilate
A:Pathway: phycoalexin biosynthesis
C:Keywords: acyltransferase; coenzyme A

Query Match 87.5%; Score 35; DB 2; Length 445;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DFGWG 5
Db 392 DFGWG 396

RESULT 12
T10711
anthranilate N-benzoyltransferase (EC 2.3.1.144) - clove pink
N:Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase
C:Species: Dianthus caryophyllus (clove pink)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C:Accession: T10711
R:Yang, Q.; Matern, U.; Grtmlyg, B.

Plant Mol. Biol. 38, 1201-1214, 1998
A:Title: Anthranilate N-hydroxycinnamoyl/benzoyltransferase gene from carnation: Rapid
A:Reference number: z17093; MUID:99084770
A:Accession: T10711
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-446 <YAN>
A:Cross-references: EMBL:Z98758; NID:e1309511; PID:e1309512
C:Function:
A:Description: catalyzes the synthesis of anthranilate
A:Pathway: phytoalexin biosynthesis
C:Keywords: acyltransferase; coenzyme A

Query Match 87.5%; Score 35; DB 2; Length 446;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWG 5
|||||
DB 393 DFGWG 397

RESULT 13
T33865
Hypothetical protein H04M03.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33865
R:Edwards, J.; Schaller, M.
Submitted to the EMBL Data Library, February 1999
A:Description: The sequence of C. elegans cosmid H04M03.
A:Reference number: Z21425
A:Accession: T33865
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-485 <EDM>
A:Cross-references: EMBL:AF154442; PIDN:AMD12787.1; GSPDB:GN00022; CESP:H04M03.4
A:Experimental source: strain Bristol N2; clone H04M03
C:Genetics:
A:Gene: CESP:H04M03.4
A:Map position: 4
A:Introns: 40/3; 137/3; 258/3; 312/3; 376/3; 419/3

Query Match 87.5%; Score 35; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWG 5
|||||
DB 211 DFGWG 215

RESULT 14
T01056
Hypothetical protein YUP8H12R.39 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 22-Oct-1999
C:Accession: T01056
R:Theologis, A.; Vysotskaya, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; Kwar
Oefner, P.; Davis, R.W.
Submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.
A:Reference number: Z14227
A:Accession: T01056
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-572 <THE>
A:Cross-references: EMBL:NC002986; NID:g2494106; PID:g3152598; GSPDB:GN00059; ATSP:YUP8H
C:Genetics:
A:Gene: ATSP:YUP8H12R.39
A:Map position: 1

A:Introns: 71/2; 98/1; 257/3

Query Match 87.5%; Score 35; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWG 5
|||||
DB 506 DFGWG 510

RESULT 15
S30515
Wound-induced protein - western balsam poplar
C:Species: Populus trichocarpa (western balsam poplar)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C:Accession: S30515; S21011
R:Perkins, E.J.; Gordon, M.P.
Submitted to the EMBL Data Library, November 1990
A:Description: Accumulation of two wound-responsive mRNAs in a poplar hybrid.
A:Reference number: S30515
A:Accession: S30515
A:Molecule type: mRNA
A:Residues: 1-113 <PER>
A:Cross-references: EMBL:X56752; NID:g20964; PID:g20965; EMBL:X55440; NID:g20955; PID

Query Match 85.0%; Score 34; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 FCGWK 6
|||||
DB 53 FCGWK 57

Search completed: October 20, 2000, 01:36:19
Job time: 117 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2000, 01:35:12 ; Search time 8.4 Seconds
(Without alignments)
22.820 Million cell updates/sec

Title: US-08-894-356c-21
Perfect score: 40
Sequence: 1 DRGWK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues
Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	87.5	918	1	HXK1_BOVIN
2	35	87.5	1045	1	Y052_HUMAN
3	34	85.0	129	1	YN09_YEAST
4	34	85.0	519	1	TKTC_CRAPL
5	34	85.0	667	1	TKTC_BACSU
6	32	80.0	298	1	VG17_HSV1
7	31	77.5	106	1	Y402_METYA
8	31	77.5	218	1	RGFE_MOUSE
9	31	77.5	273	1	OPSR_CANPA
10	31	77.5	291	1	BHCL1_RHOGE
11	31	77.5	292	1	BPHC_PSEPI
12	31	77.5	299	1	BPHC_PSEPA
13	31	77.5	393	1	HEMX_ECOLI
14	31	77.5	484	1	GLGA_BACSU
15	31	77.5	747	1	MEPA_MOUSE
16	31	77.5	748	1	MEPA_RAT
17	31	77.5	922	1	NRPI_RAT
18	31	77.5	923	1	NRPI_HUMAN
19	31	77.5	923	1	NRPI_MOUSE
20	31	77.5	1045	1	PRIS_SERMA
21	31	77.5	1045	1	PRIS_SERMA
22	30	75.0	236	1	RIP3_SAPOF
23	30	75.0	292	1	RIP2_SAPOF
24	30	75.0	293	1	RIP0_DIAOA
25	30	75.0	299	1	RIP6_SAPOF
26	30	75.0	337	1	STCP_STRPY
27	30	75.0	357	1	SH5A_HUMAN
28	30	75.0	357	1	SH5A_MOUSE
29	30	75.0	357	1	SH5A_RAT
30	30	75.0	370	1	SH5B_MOUSE
31	30	75.0	370	1	SH5B_RAT
32	30	75.0	420	1	NAP1_HUMAN
33	30	75.0	451	1	NTP1_ENTHR

ALIGNMENTS

RESULT	ID	STANDARD	PRT	918 AA
1	HXK1_BOVIN			
AC	P27595			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	HEXOKINASE, TYPE I (EC 2.7.1.1) (HK I) (BRAIN FORM HEXOKINASE).			
GN	HKL			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 89247146.			
RA	Griffin L.D., Macgregor G.R., Muzny D.M., Harter J., Cook R.G.,			
RA	McCabe E.R.;			
RT	"Synthesis and characterization of a bovine hexokinase 1 cDNA probe			
RT	by mixed oligonucleotide primed amplification of cDNA using high			
RT	complexity primer mixtures."			
RL	Biochem. Med. Metab. Biol. 41:125-131(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 92147096.			
RA	Griffin L.D., Geib B.D., Wheeler D.A., Davison D., Adams V.,			
RA	McCabe E.R.;			
RT	"Mammalian hexokinase 1: evolutionary conservation and structure to			
RT	function analysis."			
RL	Genomics 11:1014-1024(1991).			
CC	-1- CATALYTIC ACTIVITY: ATP + D-HEXOSE = ADP + D-HEXOSE 6-PHOSPHATE.			
CC	-1- ENZYME REGULATION: HEXOKINASE IS AN ALLOSTERIC ENZYME INHIBITED			
CC	BY ITS PRODUCT GLC-6-P.			
CC	-1- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.			
CC	-1- SUBUNIT: MONOMER.			
CC	-1- SUBCELLULAR LOCATION: BOUND TO THE OUTER MITOCHONDRIAL MEMBRANE.			
CC	-1- ITS HYDROPHOBIC N-TERMINAL SEQUENCE MAY BE INVOLVED IN MEMBRANE			
CC	BINDING.			
CC	-1- MISCELLANEOUS: IN VERTEBRATES THERE ARE FOUR MAJOR GLUCOSE-			
CC	PHOSPHORYLATING ISOENZYMES, DESIGNATED HEXOKINASE I, II, III AND			
CC	IV (GLUCOKINASE).			
CC	-1- SIMILARITY: THE N- AND C-TERMINAL HALVES OF THIS HEXOKINASE			
CC	SHOW EXTENSIVE SEQUENCE SIMILARITY TO EACH OTHER. THE CATALYTIC			
CC	ACTIVITY IS ASSOCIATED WITH THE C-TERMINUS WHILE REGULATORY			
CC	FUNCTION IS ASSOCIATED WITH THE N-TERMINUS.			
CC	-1- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M65140; AAA51661.1; -.			

```

DR HSSP: P19367; 1HKB.
DR INTERPRO: IPR001312; -.
DR PFAM: PF00349; hexokinase; 2.
DR PRINTS: PR00475; HEXOKINASE.
DR PROSITE: PS00378; HEXOKINASES; 2.
KM Transferase; Kinase; Glycolysis; Allosteric enzyme; Duplication;
KW ATP-binding; Membrane.
FT DOMAIN 1 12
FT DOMAIN 13 475
FT DOMAIN 149 918
FT DOMAIN 175 175
FT DOMAIN 597 623
FT NP_BIND 84 89
FT NP_BIND 532 537
FT BINDING 558 558
SO SEQUENCE 918 AA; 103064 MW; 1DCEBF7E1D06FE2B6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 35; DB 1; Length 918;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWG 5
DB 246 DFGWG 250

RESULT 2
Y052_HUMAN
ID Y052_HUMAN STANDARD; PRT; 1045 AA.
AC P42285;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0052 (FRAGMENT).
GN KIAA0052.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RX MEDLINE; 96051398.
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarayashi Y., Ishikawa K.-I., Tabata S.;
RT Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 1:223-229(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SKI2 SUBFAMILY OF HELICASES.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC INTRODUCED IN POSITION 837 TO MAKE THE PROTEIN LONGER AND MAXIMIZE
CC THE SIMILARITY WITH SKI2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D29641; BAA06124.1; ALT_FRAME.
DR INTERPRO: IPR001410; -.
DR INTERPRO: IPR001650; -.
DR PFAM: PF00270; DEAD; 1.
DR PFAM: PF00271; helicase_C; 1.
KW Hypothetical protein; Helicase; ATP-binding; Nuclear protein.
FT NON_TER 1
FT NP_BIND 165 172
FT SITE 256 259
FT SEQUENCE 1045 AA; 118243 MW; 7B16FF8E78049C20 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 35; DB 1; Length 1045;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWG 5
DB 672 DFGWG 676

RESULT 3
YNO9_YEAST
ID YNO9_YEAST STANDARD; PRT; 129 AA.
AC P53903;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 15.1 KDA PROTEIN IN RPC8-MFA2 INTERGENIC REGION.
GN YNL149C OR N1774.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE; 96287653.
RA Nasr F., Becam A.-M., Herbert C.J.;
RT "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
RT 24 complete open reading frames: 18 correspond to new genes, one of
RT which encodes a protein similar to the human myotonic dystrophy
RT kinase."
RL Yeast 11:169-175(1996).
RN [2]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN-S288C;
RX MEDLINE; 96109932.
RA Mallet L., Bussiereau F., Jacquet M.;
RT "A 4.5 kb segment of yeast chromosome XIV, which contains MFA2,
RT MRP2, CAP/SPV2, NAM9, FRB1/PPRI/RBP1, MOM22 and CPT1, predicts an
RT adenosine deaminase gene and 14 new open reading frames."
RL Yeast 11:1195-1209(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X92517; CAA63290.1; -.
DR EMBL: 271426; CAA86037.1; -.
DR EMBL: 271424; CAA96033.1; -.
DR SGD: S0005093; YNL149C.
KW Hypothetical protein.
SO SEQUENCE 129 AA; 15053 MW; A1A41E3DA8CEAA06 CRC64;

Query Match
Best Local Similarity 100.0%; Score 34; DB 1; Length 129;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 FGVGK 6
DB 84 FGVGK 88

RESULT 4
TKTC_CRAPL
ID TKTC_CRAPL STANDARD; PRT; 519 AA.
AC Q42676;
DT 01-NOV-1997 (Rel. 35, Created)

```

DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 01-NOV-1997 (Rel. 35, last annotation update)
DE TRANSEPTOLASE, CHLOROPLAST (EC 2.2.1.1) (TK) (FRAGMENT).
GN TKT3.
OS Craterostigma plantagineum.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Lamiales; Scrophulariaceae; Craterostigma.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 95163594.
RA Bernacchia G., Schwall G., Lottspeich F., Salamini F., Bartels D.;
RT "The transeptolase gene family of the resurrection plant
RT Craterostigma plantagineum: differential expression during the
RT rehydration phase."
RL EMBL J. 14:610-618(1995).
CC -1- CATALYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE + D-GLYCERALDEHYDE
CC 3-PHOSPHATE -> D-RIBOSE 5-PHOSPHATE + D-XYLULOSE 5-PHOSPHATE.
CC -1- COFACTOR: THIAMINE PYROPHOSPHATE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (POTENTIAL).
CC -1- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN LEAVES AND ROOTS.
CC -1- SIMILARITY: BELONGS TO THE TRANSEPTOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: 246646; CAA86607.1; -.
DR HSSP: P23254; 1TKA.
DR INTERPRO: IPR000360; -.
DR PFAM: PF00456; transeptolase: 1.
DR PROSITE: PS00801; TRANSEPTOLASE_1; PARTIAL.
DR PROSITE: PS00802; TRANSEPTOLASE_2; 1.
KM transeptase; Thiamine pyrophosphate; Chloroplast; Multigene family.
FT NON_TER 1
SQ SEQUENCE 519 AA; 56187 MW; 7AFD034CE2EC5685 CRC64;

OY 2 FGNGK 6
Db 473 FGNGK 477

Query Match 85.0%; Score 34; DB 1; Length 519;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE + D-GLYCERALDEHYDE
CC 3-PHOSPHATE -> D-RIBOSE 5-PHOSPHATE + D-XYLULOSE 5-PHOSPHATE.
CC -1- COFACTOR: THIAMINE PYROPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE TRANSEPTOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: 273234; CAA97616.1; -.
DR EMBL: X87845; CAA61113.1; -.
DR EMBL: 299113; CAB13673.1; -.
DR HSSP: P23254; 1AYO.
DR SUBTLIST: BG11247; TKT.
DR INTERPRO: IPR000360; -.
DR PFAM: PF00456; transeptolase: 1.
DR PROSITE: PS00801; TRANSEPTOLASE_1; 1.
DR PROSITE: PS00802; TRANSEPTOLASE_2; 1.
KM transeptase; Thiamine pyrophosphate.
SQ SEQUENCE 667 AA; 72344 MW; D93BCACD246148AF CRC64;

OY 2 FGNGK 6
Db 619 FGNGK 623

Query Match 85.0%; Score 34; DB 1; Length 667;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6
ID VG17_HSV11 STANDARD; PRT; 298 AA.
AC 000116;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, last sequence update)
DT 01-DEC-1992 (Rel. 24, last annotation update)
DE HYPOTHETICAL GENE 17 PROTEIN.
GN 17.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC unclassified Herpesviridae.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-AUBURN 1;
RX MEDLINE: 92087490.
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus."
RL Virology 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: M75136; AAA88120.1; -.
DR PIR: I36787; I36787.
KM Hypothetical protein.
SQ SEQUENCE 298 AA; 32399 MW; 4C90B1DEEA203FA9 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 298;
Best Local Similarity 66.7%; Pred. No. 80;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

OY      1 DEWGK 6
       :||||
Db      67 NFGMR 72

RESULT  7
Y402_METJA  STANDARD:  PRT:  106 AA.
ID AC O57845;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL PROTEIN M00402.
GN M00402.
OS Methanococcus jannaschli.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae:
CC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE: 96337999.
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschli."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U67492; AAB98395.1; -.
DR TIGR: M00402; -.
KW Hypothetical protein.
SQ SEQUENCE 106 AA; 12600 MW; 6398888ABD46CF3 CRC64;

Query Match      77.5%; Score 31; DB 1; Length 106;
Best Local Similarity 80.0%; Pred. No. 48;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      2 FGWKG 6
       :||||
Db      2 FGWGR 6

RESULT  8
FGGF_MOUSE  STANDARD:  PRT:  218 AA.
ID AC O35622;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FIBROBLAST GROWTH FACTOR-15 PRECURSOR (FGF-15).
GN FGF15.
OS Mus musculus (Mouse).
OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi;
OC Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97454242.
RA McWhirter J.R., Goulding M., Welner J.A., Chun J., Murre C.;

```

```

RT "A novel fibroblast growth factor gene expressed in the developing
RT nervous system is a downstream target of the chimeric homeodomain
RT oncoprotein E2A-Pbx1."
RL Development 124:3221-3232(1997).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN REGULATING CELL DIVISION
CC AND PATTERNING WITHIN SPECIFIC REGIONS OF THE EMBRYONIC BRAIN,
CC SPINAL CORD AND SENSORY ORGANS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING BRAIN.
CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF007268; AAB63197.1; -.
DR MGD: MGI:1096383; FGF15.
DR INTERPRO: IPR002209; -.
DR INTERPRO: IPR002348; -.
DR PFAM: PF00167; FGF; 1.
DR PRINTS: PR00262; IL1HBGF.
DR PROSITE: PS00247; HBGF_FGF; 1.
KW Growth factor; Signal.
FT SIGNAL 1 25
FT CHAIN 26 218 POTENTIAL.
FT SEQUENCE 218 AA; 25236 MW; A96B0D771FE125A5 CRC64;

Query Match      77.5%; Score 31; DB 1; Length 218;
Best Local Similarity 80.0%; Pred. No. 90;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      2 FGWKG 6
       :||||
Db      44 YGWKG 48

RESULT  9
OPSR_CANFA  STANDARD:  PRT:  273 AA.
ID AC O18914;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 38, Last annotation update)
DE RED-SENSITIVE OPSIN (RED CONE PHOTORECEPTOR PIGMENT) (FRAGMENT).
GN RCP.
OS Canis familiaris (Dog).
OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi;
OC Mammalia: Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98242086.
RA Yokoyama S., Radlwimmer F.B.;
RT "The 'five-sites' rule and the evolution of red and green color
RL vision in mammals."
RL Mol. Biol. Evol. 15:560-567(1998).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: THE THREE COLOR PIGMENTS ARE FOUND IN THE CONE
CC PHOTORECEPTOR CELLS.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF031533; AAB86633.1; .
 DR GCRDB: GCR_2500; .
 DR INTERPRO: IPR000276; .
 DR INTERPRO: IPR001760; .
 DR PIRAM: PF00001; 7tm_1; 1.
 DR PROSITE: PS00237; G-PROTEIN-RECEPTOR; 1.
 DR PROSITE: PS00238; OPSIN; PARTIAL.
 KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 KW phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
 FT NON_TER 1 1
 FT DOMAIN 1 1
 FT TRANSSEM 6 5
 FT DOMAIN 31 30
 FT TRANSSEM 43 42
 FT TRANSSEM 69 68
 FT DOMAIN 83 82
 FT TRANSSEM 103 102
 FT DOMAIN 122 121
 FT TRANSSEM 146 145
 FT TRANSSEM 172 171
 FT TRANSSEM 199 198
 FT DOMAIN 200 221
 FT TRANSSEM 222 245
 FT DOMAIN 246 253
 FT TRANSSEM 254 273
 FT DISULFID 79 156
 FT BINDING 265 265
 FT NON_TER 273 273
 SO SEQUENCE 273 AA; 30373 MW; C1A27C95D1649F85 CRC64;
 Query Match 77.5%; Score 31; DB 1; Length 273;
 Best Local Similarity 80.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 FGWGR 6
 Db 142 FGWGR 146
 RESULT 10
 BHCL_RHOG
 ID BHCL_RHOG STANDARD; PRT; 291 AA.
 AC P47231;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BIPEHYL-2,3-DIOL 1,2-DIOXYGENASE I (EC 1.13.11.39) (23OHP
 DE OXYGENASE I) (2,3-DIHYDROXYBIPEHYL DIOXYGENASE I) (DHBD I).
 GN BPHCL.
 OS Rhodococcus globerulus.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.
 RN [1]
 RC STRAIN-P6.
 RX MEDLINE; 94171820.
 RA Asturias J.A., Ellis L.D., Prucha M., Timmls K.N.;
 RA "Analysis of three 2,3-dihydroxydiphenyl 1,2-dioxygenases found in
 RA Rhodococcus globerulus P6. Identification of a new family of
 RA extradiol dioxygenases.";
 RL J. Biol. Chem. 269:7807-7815(1994).
 CC -1- CATALYTIC ACTIVITY: BIPEHYL-2,3-DIOL + O(2) - 2-HYDROXY-6-OXO-
 CC 6-PHENYLHEXA-2,4-DIENATE + H(2)O.
 CC -1- Cofactor: FERROUS ION.
 CC -1- PATHWAY: DEGRADATION OF BIPEHYLS AND POLYCHLOROBIPHENYLS (PCB) TO
 CC BENZOIC ACID AND CHLOROBENZOIC ACIDS.
 CC -1- SIMILARITY: BELONGS TO THE EXTRADIOL RING-CLEAVAGE DIOXYGENASE

CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X75633; CA53297.1; .
 DR HSSP: P47228; 1HAN.
 DR INTERPRO: IPR000486; .
 DR PIRAM: PF01033; Extradiol_dioxy; 1.
 DR PROSITE: PS00082; EXTRADIOL_DIOXYGENAS; 1.
 KW Oxidoreductase; Dioxygenase; Aromatic hydrocarbons catabolism; Iron.
 FT METAL 146 146
 FT METAL 210 210
 FT METAL 260 260
 FT METAL 260 260
 SO SEQUENCE 291 AA; 32081 MW; 104F189FEDD6A CRC64;
 Query Match 77.5%; Score 31; DB 1; Length 291;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DFGWG 5
 Db 260 DFGWG 264
 RESULT 11
 BPHC_PSES1
 ID BPHC_PSES1 STANDARD; PRT; 292 AA.
 AC P17297; Q52441;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BIPEHYL-2,3-DIOL 1,2-DIOXYGENASE (EC 1.13.11.39) (23OHP OXYGENASE)
 DE (2,3-DIHYDROXYBIPEHYL DIOXYGENASE) (DHBD).
 GN BPHC.
 OS Pseudomonas sp. (strain KKS102).
 OC Bacteria; Proteobacteria.
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE; 89213965.
 RA Kimbara K., Hashimoto T., Fukuda M., Koana T., Takagi M., Oishi M.,
 RA Yano K.;
 RA "Cloning and sequencing of two tandem genes involved in degradation
 RA of 2,3-dihydroxydiphenyl to benzoic acid in the polychlorinated
 RA biophenyl-degrading soil bacterium Pseudomonas sp. strain KKS102.";
 RL J. Bacteriol. 171:2740-2747(1989).
 GN [2]
 GN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE; 96226036.
 RA Senda T., Sugiyama K., Narita H., Yamamoto T., Kimbara K., Fukuda M.,
 RA Sato M., Yano K., Mitsui Y.;
 RA "Three-dimensional structures of free form and two substrate
 RA complexes of an extradiol ring-cleavage type dioxygenase, the BphC
 RA enzyme from Pseudomonas sp. strain KKS102.";
 RL J. Mol. Biol. 255:735-752(1996).
 CC -1- CATALYTIC ACTIVITY: BIPEHYL-2,3-DIOL + O(2) - 2-HYDROXY-6-OXO-
 CC 6-PHENYLHEXA-2,4-DIENATE + H(2)O.
 CC -1- Cofactor: FERROUS ION.
 CC -1- PATHWAY: DEGRADATION OF BIPEHYLS AND POLYCHLOROBIPHENYLS (PCB) TO
 CC BENZOIC ACID AND CHLOROBENZOIC ACIDS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE EXTRADIOL RING-CLEAVAGE DIOXYGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M26433; AAA25750.1; -
 DR EMBL: D17319; BAA04141.1; -
 DR PIR: A32312; DAPSPC.
 DR PDB: 1DHV; 15-OCT-95.
 DR INTERPRO: IPR000486; -
 DR PFAM: PF01013; Extradiol_dioxy; 1.
 DR PROSITE: PS00082; EXTRADIOL_DIOXYGENAS; 1.
 DR Oxidoreductase; Dioxygenase; Aromatic hydrocarbons catabolism; Iron;
 KW 3D-structure.
 FT INIT_MET 0 0
 FT METAL 145 145 IRON.
 FT METAL 209 209 IRON.
 FT METAL 260 260 IRON.
 SQ SEQUENCE 292 AA; 32113 MW; 5D7D912F79EA8476 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 292;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWG 5
 :||||
 Db 260 EFGWG 264

RESULT 12
 BPHC_PSEPA STANDARD; PRT; 299 AA.
 ID BPHC_PSEPA P11122;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BIPIHENYL-2,3-DIOL 1,2-DIOXYGENASE (EC 1.13.11.39) (23OHP OXYGENASE)
 DE (2,3-DIHYDROXYBIPIHENYL DIOXYGENASE) (DHBD).
 GN BPHC.
 OS Pseudomonas paucimobillis (Sphingomonas paucimobillis).
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
 OC Sphingomonas.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Q1;
 RX MEDLINE: 88326911.
 RA Teira K., Hayase N., Arimura N., Yamashita S., Miyazaki T.,
 Furukawa K.;
 RT "Cloning and nucleotide sequence of the 2,3-dihydroxybiphenyl
 dioxygenase gene from the PCB-degrading strain of Pseudomonas
 paucimobillis Q1.";
 RT Biochemistry 27:3990-3996(1988).
 CC -1- CATALYTIC ACTIVITY: BIPIHENYL-2,3-DIOL + O(2) = 2-HYDROXY-6-EXO-
 6-PHENYLHEXA-2,4-DIENATE + H(2)O.
 CC -1- COFACTOR: FERROUS ION.
 CC -1- PATHWAY: DEGRADATION OF BIPIHENYLS AND POLYCHLOROBIPHENYLS (PCB) TO
 BENZOIC ACID AND CHLOROBENZOIC ACIDS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE EXTRADIOL RING-CLEAVAGE DIOXYGENASE
 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M20640; AAA25678.1; -
 DR PIR: A28718; A28718.
 DR HSSP: P17297; 1DHV.

DR INTERPRO: IPR000486; -
 DR PFAM: PF01013; Extradiol_dioxy; 1.
 DR PROSITE: PS00082; EXTRADIOL_DIOXYGENAS; 1.
 DR Oxidoreductase; Dioxygenase; Aromatic hydrocarbons catabolism; Iron.
 KW 149 149 IRON (BY SIMILARITY).
 FT METAL 149 149 IRON (BY SIMILARITY).
 FT METAL 212 212 IRON (BY SIMILARITY).
 FT METAL 263 263 IRON (BY SIMILARITY).
 SQ SEQUENCE 299 AA; 33095 MW; 29746CAB79D6FC7B CRC64;

Query Match 77.5%; Score 31; DB 1; Length 299;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWG 5
 :||||
 Db 263 EFGWG 267

RESULT 13
 HEMX_ECOLI STANDARD; PRT; 393 AA.
 ID HEMX_ECOLI P09127;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PUTATIVE UROPORPHYRIN-III C-METHYLTRANSFERASE (EC 2.1.1.107) (UROGEN
 DE III METHYLASE) (ORF X).
 GN HEMX.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 89096348.
 RA Sasaman A., Echelard Y., Letowski J., Tardif D., Drolet M.;
 RT "Nucleotide sequence of the hemx gene, the third member of the uro
 operon of Escherichia coli K12.";
 RT Nucleic Acids Res. 16:11835-11835(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / CS520;
 RX MEDLINE: 89041586.
 RA Alefounder P.R.;
 RT "The sequence of hemc, hemd and two additional E. coli genes.";
 RT Nucleic Acids Res. 16:9871-9871(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 92358234.
 RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
 RT "Analysis of the Escherichia coli genome: DNA sequence of the region
 from 84.5 to 86.5 minutes.";
 RT Science 257:771-778(1992).
 RN [4]
 RP SEQUENCE OF 1-12.
 RC STRAIN-K12 / EMG2;
 RX MEDLINE: 97443975.
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12.";
 RT Electrophoresis 18:1259-1313(1997).
 CC -1- CATALYTIC ACTIVITY: 2 S-ADENOSYL-L-METHIONINE + UROPORPHYRIN III
 = 2 S-ADENOSYL-L-HOMOCYSTEINE + STROPHOMINE AND COBALAMIN.
 CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF STROHEME AND COBALAMIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@lsb-sib.ch).

CC -----

DR EMBL: X13406; CAA31772.1; -

DR EMBL: X12614; CAA31134.1; -

DR EMBL: M87049; AAC67599.1; -

DR EMBL: AE000456; AAC76806.1; -

DR PIR: S02185; S02185.

DR PIR: S30693; S30693.

DR ECODBASE: B043.0; 6TH EDITION.

DR ECODBASE: B043.1; 6TH EDITION.

DR ECGENE: EG10433; HEMX.

KW Porphylin biosynthesis; transferase; methyltransferase.

SW SEQUENCE 393 AA; 42963 MW; 9D272C6401D0E354 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 393;

Best Local Similarity 80.0%; Pred. No. 1.5e+02;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 FGWCK 6

Db 55 YGWCK 59

RESULT 14

GLGA_BACSU STANDARD; PRT; 484 AA.

ID GLGA_BACSU

AC P39125:

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE GLYCOGEN SYNTHASE (EC 2.4.1.21) (STARCH [BACTERIAL GLYCOGEN] SYNTHASE).

GN GLGA.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/staphylococcus group; Bacillus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-168;

RA MEDLINE: 94195107.

RA Kiel J.A.K.W., Boels J.M., Beldman G., Venema G.;

RT "Glycogen in Bacillus subtilis: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation.";

RT Mol. Microbiol. 11:203-218(1994).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 98048467.

RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;

RT "Sequencing and functional annotation of the Bacillus subtilis genes in the 200 kb rnb-dnaB region.";

RT Microbiology 143:3431-3441(1997).

CC -1- FUNCTION: SYNTHESIZES ALPHA-1,4-GLUCAN CHAINS USING ADP-GLUCOSE.

CC -1- CATALYTIC ACTIVITY: ADP-GLUCOSE + (1,4-ALPHA-D-GLUCOSYL)[N] - ADP + (1,4-ALPHA-D-GLUCOSYL)[N+1].

CC -1- PATHWAY: SECOND STEP IN GLYCOGEN BIOSYNTHESIS.

CC -1- INDICATION: EXPRESSED EXCLUSIVELY ON MEDIA CONTAINING CARBON SOURCES THAT ALLOW EFFICIENT SPOULATION.

CC -1- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.lsb-sib.ch/announce/> or send an email to license@lsb-sib.ch).

CC -----

DR EMBL: Z25795; CAA81043.1; -

DR EMBL: AF008220; AAC00217.1; -

DR EMBL: Z29119; CAB15073.1; -

DR PIR: S36627; S36627.

DR SUBTILIST; BG10910; GLGA.

KW Glycogen biosynthesis; transferase; glycosyltransferase.

FT BINDING 15 ADP-GLUCOSE (BY SIMILARITY).

SW SEQUENCE 484 AA; 55857 MW; D479E8064818E0E4 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 484;

Best Local Similarity 66.7%; Pred. No. 1.8e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 DEGWCK 6

Db 458 DYSGWK 463

RESULT 15

MEPA_MOUSE STANDARD; PRT; 747 AA.

ID MEPA_MOUSE

AC P28825:

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE MEPRIN A ALPHA-SUBUNIT PRECURSOR (EC 3.4.24.18) (ENDOPEPTIDASE-2) (MEP-1).

GN MEPA.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-C57BL/6, AND C3H/HE; TISSUE-KIDNEY;

RX MEDLINE: 92250517.

RA Jiang W., Gorbac C.M., Flannery A.V., Beynon R.J., Grant G.A., Bond J.S.;

RT "The alpha subunit of meprin A. Molecular cloning and sequencing, differential expression in inbred mouse strains, and evidence for divergent evolution of the alpha and beta subunits.";

RT J. Biol. Chem. 267:9185-9193(1992).

RN [2]

RP SEQUENCE OF 64-247 FROM N.A.

RX MEDLINE: 92042028.

RA Dumermuth E., Sterchi E.E., Jiang W., Wolz R.L., Bond J.S., Flannery A.V., Beynon R.J.;

RT "The astacin family of metalloendopeptidases.";

RT J. Biol. Chem. 266:21381-21385(1991).

RN [3]

RP SEQUENCE OF 62-258 FROM N.A.

RC STRAIN-129;

RA MEDLINE: 97305147.

RA Jiang W., Flannery A.V.;

RT "Correlation of the exon/intron organization to the secondary structures of the protease domain of mouse meprin alpha subunit.";

RT Gene 189:65-71(1997).

RN [4]

RP CHARACTERIZATION.

RX MEDLINE: 91355206.

RA Wolz R.L., Harris R.B., Bond J.S.;

RT "Mapping the active site of meprin-A with peptide substrates and inhibitors.";

RT Biochemistry 30:8488-8493(1991).

CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEIN AND PEPTIDE SUBSTRATES PREFERENTIALLY ON CARBOXYL SIDE OF HYDROPHOBIC RESIDUES.

CC -1- COFACTOR: BINDS ONE ZINC ION.

CC -1- SUBUNIT: HOMOTETRAMER OF ALPHA OR BETA SUBUNITS; HETEROTETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS ARE FORMED BY NON-COVALENT ASSOCIATION OF TWO DISULFIDE-LINKED HETERODIMERS; GENETIC FACTORS DETERMINE WHICH OLIGOMER(S) WILL BE FORMED (STRAIN-SPECIFIC).

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: KIDNEY, INTESTINAL BRUSH BORDERS, AND SALIVARY DUCTS.

CC -1- PTM: N-GLYCOSYLATED; AT LEAST 3 OF THE POTENTIAL SITES ARE USED.

CC -1- SIMILARITY: CONTAINS 1 BGF-LIKE DOMAIN.

CC -! SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC
CC METALLOPROTEASE). ALSO KNOWN AS THE ASTACIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74897; AAA75354.1; ALT_INIT.
DR EMBL: U62765; AAC53194.1; -.
DR PIR: A40195; A40195.
DR PDB: 1IAF; 31-AUG-94.
DR MCD; MGI:96963; MEPIA.
DR INTERPRO; IPR000130; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000998; -.
DR INTERPRO; IPR001506; -.
DR INTERPRO; IPR002083; -.
DR PFAM; PF01400; Astactin; 1.
DR PFAM; PF00008; EGF; 1.
DR PFAM; PF00629; MAM; 1.
DR PFAM; PF00917; MATH; 1.
DR PRINTS; PRO00020; MAMDOMAIN.
DR PRINTS; PRO0480; ASTACTIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS50060; MAM_2; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
KM Hydrolase: Metalloprotease: Zinc: Glycoprotein; Transmembrane;
KV Zymogen: Signal; EGF-like domain; 3D-structure.
FT SIGNAL 1 20
FT PROPEP 1 64
FT CHAIN 65 747
FT DOMAIN 65 713
FT TRANSMEM 714 741
FT DOMAIN 742 747
FT DOMAIN 65 262
FT DOMAIN 263 432
FT DOMAIN 671 711
FT METAL 154 154
FT METAL 155 155
FT ACT SITE 158 158
FT METAL 164 164
FT DISULFD 675 686
FT DISULFD 680 695
FT DISULFD 697 710
FT CARBOHYD 28 28
FT CARBOHYD 139 139
FT CARBOHYD 221 221
FT CARBOHYD 257 257
FT CARBOHYD 317 317
FT CARBOHYD 413 413
FT CARBOHYD 439 439
FT CARBOHYD 533 533
FT CARBOHYD 540 540
FT CARBOHYD 601 601
FT CONFLICT 259 259
SQ SEQUENCE 747 AA; 84197 MM; 2A18242AA3505633 CRR64;

Query Match 77.5%; Score 31; DB 1; Length 747;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Search completed: October 20, 2000, 01:37:07
Job time: 115 sec

OY 1 DRGNK 6
DB 565 DWGMGO 570

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2000, 01:33:12 ; Search time 14.51 Seconds
(without alignments)
14.139 Million cell updates/sec

Title: US-08-894-356c-21

Perfect score: 40
Sequence: 1 DFGWGK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq-36:*

- 1: /SID56/gcgdata/geneseq/AA1980.DAT:*
- 2: /SID56/gcgdata/geneseq/AA1981.DAT:*
- 3: /SID56/gcgdata/geneseq/AA1982.DAT:*
- 4: /SID56/gcgdata/geneseq/AA1983.DAT:*
- 5: /SID56/gcgdata/geneseq/AA1984.DAT:*
- 6: /SID56/gcgdata/geneseq/AA1985.DAT:*
- 7: /SID56/gcgdata/geneseq/AA1986.DAT:*
- 8: /SID56/gcgdata/geneseq/AA1987.DAT:*
- 9: /SID56/gcgdata/geneseq/AA1988.DAT:*
- 10: /SID56/gcgdata/geneseq/AA1989.DAT:*
- 11: /SID56/gcgdata/geneseq/AA1990.DAT:*
- 12: /SID56/gcgdata/geneseq/AA1991.DAT:*
- 13: /SID56/gcgdata/geneseq/AA1992.DAT:*
- 14: /SID56/gcgdata/geneseq/AA1993.DAT:*
- 15: /SID56/gcgdata/geneseq/AA1994.DAT:*
- 16: /SID56/gcgdata/geneseq/AA1995.DAT:*
- 17: /SID56/gcgdata/geneseq/AA1996.DAT:*
- 18: /SID56/gcgdata/geneseq/AA1997.DAT:*
- 19: /SID56/gcgdata/geneseq/AA1998.DAT:*
- 20: /SID56/gcgdata/geneseq/AA1999.DAT:*
- 21: /SID56/gcgdata/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	6	17	W04728
2	40	100.0	439	21	Y51247
3	40	100.0	446	17	W04725
4	40	100.0	448	15	R47475
5	40	100.0	448	15	R46929
6	40	100.0	448	17	W04724
7	40	100.0	450	17	W04727
8	40	100.0	454	17	W04726
9	40	100.0	459	17	W04722
10	40	100.0	479	17	W04723
11	37	92.5	10	21	Y51235
12	35	87.5	391	21	Y68644

13	35	87.5	391	21	Y68655
14	35	87.5	433	20	Y17400
15	35	87.5	438	20	Y17401
16	35	87.5	439	20	Y17402
17	35	87.5	439	20	Y17403
18	34	85.0	76	18	W27832
19	34	85.0	324	19	Y85934
20	34	85.0	345	21	Y68646
21	34	85.0	346	21	Y68647
22	34	85.0	361	21	Y70342
23	34	85.0	361	21	Y44735
24	34	85.0	517	21	Y69879
25	32	80.0	60	20	Y02727
26	32	80.0	331	20	W68971
27	32	80.0	845	18	W17890
28	32	80.0	845	19	W56579
29	32	80.0	1849	19	W56573
30	32	80.0	2516	18	W17899
31	32	80.0	2516	19	W56572
32	31	77.5	45	19	Y20659
33	31	77.5	111	19	W38667
34	31	77.5	116	18	W20660
35	31	77.5	128	18	W20127
36	31	77.5	213	20	Y08582
37	31	77.5	245	20	Y17873
38	31	77.5	285	21	Y81540
39	31	77.5	300	18	W18660
40	31	77.5	656	21	Y44944
41	31	77.5	921	20	W96247
42	31	77.5	922	20	W96309
43	31	77.5	923	20	Y06317
44	31	77.5	923	20	Y23247
45	31	77.5	923	20	W96246

ALIGNMENTS

RESULT 1	
W04728	W04728 standard; Protein; 6 AA.
AC	W04728;
DT	06-FEB-1997 (first entry)
DE	Aromatic acyl transferase peptide fragment.
XX	
KW	Aromatic acyl transferase; transformation; anthocyanin pigment;
KW	plants; acylation; colour; tone; colouration; colour change;
KW	Gentiana triflora; Petunia hybrida; Petilla oclimoides;
KW	Scenecio cruentus; Lavandula angustifolia.
XX	
OS	Gentiana triflora var. japonica (Clone pGAT4).
XX	
PN	W09625500-A1.
PD	22-AUG-1996.
XX	
PF	16-FEB-1996; 96MO-JP00348.
XX	
PR	30-JAN-1996; 96JP-0046534.
PR	17-FEB-1995; 95JP-0067159.
PR	29-JUN-1995; 95JP-0196915.
XX	
PA	(SUNR) SUNTORY LTD.
PI	Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
PI	Nakao M, Tanaka Y, Yonekura K;
XX	
DR	WPI; 1996-393401/39.
XX	
PT	DNA coding for aromatic acyl transferase - for transforming plants

Amino acid sequenc
Clarkia breweri be
Clarkia concinna b
Clarkia concinna b
Staphylococcus aur
S. pneumoniae deri
Amino acid sequenc
Amino acid sequenc
Human G-protein co
Human G-protein co
B. lactofermentum
Human secreted pro
Alcaligenes sp. pr
Photobacterium lum
Fragment of toxin
Toxin TcdA11, enco
Photobacterium lum
Toxin TcdA, encode
Human neurofilamen
Streptococcus pneu
H. pylori cytoplas
H. pylori cytoplas
Human GGF-15 prote
Human TRANCE. Hom
Streptococcus pneu
Fragmented human N
Wheat sulphate per
Rat semaphorin rec
Neurophilin. Ratu
Human neurophilin-1
Human VEGF165/NP-
Human semaphorin r

PT which produce anthocyanin pigments and thus altering colour tone,
 PT e.g. of flowers
 XX
 PS Claim 2; Page 81; 94pp; Japanese.
 XX
 CC Vectors containing DNA fragments encoding proteins of plant origin
 CC with aromatic acyl transferase activity may be used to transform
 CC plants which produce anthocyanin pigments. The aromatic acyl
 CC transferase acylates the pigments in the flower resulting in colour
 CC tone changes and allowing new colourations to be produced. Six
 CC specific DNA sequences encoding aromatic acyl transferase from
 CC different plants are described in T37308-T37313. This peptide
 CC fragment was isolated from three of the six clones and was used to
 CC synthesise a degenerate primer (T37314) which was then used to
 CC identify other aromatic acyl transferase encoding clones.
 XX
 SQ Sequence 6 AA;
 Query Match 100.0%; Score 40; DB 17; Length 6;
 Best Local Similarity 100.0%; Pred. No. 2.1e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DFGWKG 6
 DB 1 dfgwkg 6
 RESULT 2
 Y51247
 ID Y51247 standard; peptide: 439 AA.
 XX
 AC Y51247;
 XX
 DT 07-APR-2000 (first entry)
 XX
 DE C. roseus DAT protein.
 XX
 KW Acyltransferase; plant; RACE; screening; anti-DAT antibody; DAT;
 KW decetylvinidoline 4-O-acetyltransferase.
 XX
 OS Catharanthus roseus.
 XX
 PN CA2231621-A1.
 XX
 PD 10-SEP-1999.
 XX
 PF 10-MAR-1998; 98CA-2231621.
 XX
 PR 10-MAR-1998; 98CA-2231621.
 XX
 PA (UYMO-) UNIV MONTREAL.
 XX
 PI St-Pierre B, De Luca V;
 XX
 DR WPI; 2000-087682/08.
 XX
 PT Novel amino acid consensus sequence useful for the identification of
 PT acyltransferase enzymes in plants -
 XX
 PS Disclosure: Fig 2; 57pp; English.
 XX
 CC This invention describes a novel consensus sequence (1) for the
 CC identification of acyltransferase enzyme in plants. (1) comprises
 CC His-X₁-X₂-X₃-Asp-X₄ where X₁ - an amino acid comprising Ala, Arg,
 CC Thr, Cys, Asn, His, Met, Ile, Leu, Val or Lys; X₂ - an amino acid
 CC comprising Val, Ile, Leu or Met; X₃ - an amino acid comprising Cys,
 CC Phe, Val, Leu, Met, Ala, Gly, Ser or Thr; and X₄ - an amino acid
 CC comprising Gly, Ala, Ile, Met or Val. The invention also describes a
 CC novel method for the identification of acyltransferase enzyme in plants
 CC comprising: (a) the steps of (1) DNA amplification using PCR primers
 CC specific for (1) coupled to 3'- or 5'- RACE (rapid amplification of
 CC cDNA ends) protocols; (11) direct screening of cDNA and genomic

CC libraries using degenerate primers based on (1); (11) screening of
 CC cDNA expression libraries using antibodies specific for (1); or (iv)
 CC screening of cDNA expression libraries using anti-DAT (decetylvinidoline
 CC 4-O-acetyltransferase) antibodies; (b) identifying in clones of (1);
 CC (c) identifying homology to the DAT gene family; and (d) identifying an
 CC open reading frame encoding a protein with a molecular weight between 45
 CC and 54kDa. This sequence represents the Catharanthus roseus DAT protein
 CC which is described in the method of the invention.
 XX
 SQ Sequence 439 AA;
 Query Match 100.0%; Score 40; DB 21; Length 439;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DFGWKG 6
 DB 380 dfgwkg 385
 RESULT 3
 W04725
 ID W04725 standard; Protein; 446 AA.
 XX
 AC W04725;
 XX
 DT 06-FEB-1997 (first entry)
 XX
 DE Aromatic acyl transferase of Perilla ocimoides.
 XX
 KW Aromatic acyl transferase; transformation; anthocyanin pigment;
 KW plants; acylation; colour; tone; colouration; colour change;
 KW Gentiana triflora; Petunia hybrida; Perilla ocimoides;
 KW Senecio cruentus; Lavandula angustifolia.
 XX
 OS Perilla ocimoides (Clone PSAT208).
 XX
 PN W09625500-A1.
 XX
 PD 22-AUG-1996.
 XX
 PF 16-FEB-1996; 96WO-JP00348.
 XX
 PR 30-JAN-1996; 96JP-0046534.
 PR 17-FEB-1995; 95JP-0067159.
 PR 29-JUN-1995; 95JP-0196915.
 XX
 PA (SUNR) SUNTORY LTD.
 XX
 PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
 PI Nakao M, Tanaka Y, Yonekura K;
 XX
 DR WPI; 1996-393401/39.
 XX
 PT DNA coding for aromatic acyl transferase - for transforming plants
 PT which produce anthocyanin pigments and thus altering colour tone,
 PT e.g. of flowers
 XX
 PS Claim 4; Page 65-68; 94pp; Japanese.
 XX
 CC Vectors containing DNA fragments encoding proteins of plant origin
 CC with aromatic acyl transferase activity may be used to transform
 CC plants which produce anthocyanin pigments. The aromatic acyl
 CC transferase acylates the pigments in the flower resulting in colour
 CC tone changes and allowing new colourations to be produced. Six
 CC specific DNA sequences encoding aromatic acyl transferase from
 CC different plants are described in T37308-T37313.
 XX
 SQ Sequence 446 AA;

Query Match 100.0%; Score 40; DB 17; Length 446;
 Best Local Similarity 100.0%; Pred. No. 19;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
 |||||

Db 389 dfwgk 394

RESULT 4

R47475

ID R47475 standard; Protein; 448 AA.

AC R47475;

XX

DE 06-JUL-1994 (first entry)

Polypeptide encoded by anther specific cDNA clone ant32.

XX

KW Transgenic plants; male sterility; pollen; sterile;

KW self-pollination; seed; hybrid; toxin-A; peptide; T-uril3; Gln; 1aaL;

KW CyA; toxin; Nicotiana tabacum; Diptheria; Swinia chrysanthemi;

KW phage Mu; Pseudomonas syringae; Bacillus thuringiensis; anther;

XX breeding.

XX Nicotiana tabacum.

XX

PN EP578611-A.

XX

PD 12-JAN-1994.

XX

PF 24-JUN-1993; 93EP-0810455.

XX

PR 02-JUL-1992; 92US-0908242.

XX

PA (CIBA) CIBA GEIGY AG.

XX

PI Crossland LD, Tuttle AB;

XX

DR WPI: 1994-010428/02.

XX

DR N-PSDB; Q54685; Q53693.

XX

PT Anther-specific cDNA, genomic and recombinant DNA - produce

XX transgenic male-sterile plants, which prevents self-pollination,

PT in hybrid seed prodn.

XX

PS Disclosure: Page 25-27; 75pp; English.

XX

CC Anther specific cDNA or genomic sequences can be used to identify

CC and isolate anther specific promoters. The anther specific promoter

CC can then be cloned into a recombinant construct and used to express

CC heterologous genes. Preferred heterologous genes include Diptheria

CC toxin A-chain gene; peccate lyase gene peIF from Erwinia

CC chrysanthemi; T-uril3 from cms-T maize mitochondrial genomes; the

CC Gln recombinase gene from phage Mu; the indole acetic acid-lyase

CC synthetase gene from Pseudomonas syringae and the CyA toxin gene

CC from Bacillus thuringiensis israeliensis. All of these genes when

CC expressed in anther tissue will result in the inability of the

CC plant to produce viable pollen. Transformation of plants with such

CC a recombinant construct can produce transgenic, male sterile plants.

CC Male sterility is important in the production of hybrid seeds as it

CC prevents self pollination which hinders breeding and hybrid seed

XX production.

Sequence 448 AA;

Query Match

Best Local Similarity 100.0%; Score 40; DB 15; Length 448;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
 |||||

Db 392 dfwgk 397

RESULT 5

R46929

ID R46929 standard; Protein; 448 AA.

XX

AC R46929;

XX

DE 27-OCT-1994 (first entry)

XX

DE Tobacco Ant32 genomic clone - encoded sequence.

XX

KM Tobacco ant32 genomic protein; male-sterile plants.

XX

OS Nicotiana tabacum.

XX

PN EP589841-A.

XX

PD 30-MAR-1994.

XX

PF 15-SEP-1993; 93EP-0810654.

XX

PR 24-SEP-1992; 92US-0950348.

XX

PA (CIBA) CIBA GEIGY AG.

XX

PI Crossland LD, Stein JT, Tuttle A;

XX

DR WPI: 1994-119827/15.

XX

DR N-PSDB; Q58340.

XX

PT Dual method for producing male-sterile plants - comprises

XX crossing 2 genetically transformed plants, useful for producing

PT hybrid seed

XX

PS Example 8; Page 22-28; 30pp; English.

XX

CC R46929 is anther specific from Nicotiana tabacum.

CC The ant32 promoter can be operably linked to, eg, 77 RNA

CC polymerase gene or to GAI4/VP1 fusion gene in order to

CC produce male-sterile plants. The inventors claim that

CC male-sterile plants finally produced can be used to increase

CC hybrid vigour in plants, e.g. maize.

XX

SQ Sequence 448 AA;

Query Match

Best Local Similarity 100.0%; Score 40; DB 15; Length 448;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
 |||||

Db 392 dfwgk 397

RESULT 6

W04724

ID W04724 standard; Protein; 448 AA.

XX

AC W04724;

XX

DE 06-FEB-1997 (first entry)

Aromatic acyl transferase of Petunia hybrida.

XX

KW Aromatic acyl transferase; transformation; anthocyanin pigment;

KW plants; acylation; colour; tone; colouration; colour change;

KW Gentiana triflora; Petunia hybrida; Petilla octimoides;

KW Scenecio cruentus; Lavandula angustifolia.

XX

OS Petunia hybrida (Clone PPAR48).

DR WPI: 1996-393401/39.
 DR N-PSDB: T37312.
 XX
 PT DNA coding for aromatic acyl transferase - for transforming plants
 PT which produce anthocyanin pigments and thus altering colour tone,
 PT e.g. of flowers
 XX
 PS Claim 4; Page 69-72; 94pp; Japanese.
 XX
 CC Vectors containing DNA fragments encoding proteins of plant origin
 CC with aromatic acyl transferase activity may be used to transform
 CC plants which produce anthocyanin pigments. The aromatic acyl
 CC transferase acylates the pigments in the flower resulting in colour
 CC tone changes and allowing new colourations to be produced. Six
 CC specific DNA sequences encoding aromatic acyl transferase from
 CC different plants are described in T37308-T37313.
 XX
 SQ Sequence 454 AA;

Query Match 100.0%; Score 40; DB 17; Length 454;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 DFGMGK 6
 |||||
 Db 394 dfgwgk 399

RESULT 9
 W04722
 ID W04722 standard; Protein: 469 AA.
 XX
 AC W04722;
 XX
 DT 06-FEB-1997 (first entry)
 XX
 DE Aromatic acyl transferase of Gentiana triflora var. japonica.
 XX
 KM Aromatic acyl transferase; transformation; anthocyanin pigment;
 KM plants; acylation; colour; tone; colouration; colour change;
 KM Gentiana triflora; Petunia hybrida; Perilla octinoides;
 KM Senecio cruentus; Lavandula angustifolia.
 XX
 OS Gentiana triflora var. japonica (clone pcart4).
 XX
 PN W09625500-A1.
 XX
 PD 22-AUG-1996.
 XX
 PF 16-FEB-1996; 96WO-JP00348.
 XX
 PR 30-JAN-1996; 96JP-0046534.
 PR 17-FEB-1995; 95JP-0067159.
 PR 29-JUN-1995; 95JP-0196915.
 XX
 PA (SUNR) SUNTORY LTD.
 XX
 PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
 PI Nakao M, Tanaka Y, Yonekura K;
 XX
 DR WPI: 1996-393401/39.
 DR N-PSDB: T37308.
 XX
 PT DNA coding for aromatic acyl transferase - for transforming plants
 PT which produce anthocyanin pigments and thus altering colour tone,
 PT e.g. of flowers
 XX
 PS Claim 4; Page 53-57; 94pp; Japanese.
 XX
 CC Vectors containing DNA fragments encoding proteins of plant origin
 CC with aromatic acyl transferase activity may be used to transform
 CC plants which produce anthocyanin pigments. The aromatic acyl

CC transferase acylates the pigments in the flower resulting in colour
 CC tone changes and allowing new colourations to be produced. Six
 CC specific DNA sequences encoding aromatic acyl transferase from
 CC different plants are described in T37308-T37313.
 XX
 SQ Sequence 469 AA;

Query Match 100.0%; Score 40; DB 17; Length 469;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 DFGMGK 6
 |||||
 Db 411 dfgwgk 416

RESULT 10
 W04723
 ID W04723 standard; Protein: 479 AA.
 XX
 AC W04723;
 XX
 DT 06-FEB-1997 (first entry)
 XX
 DE Aromatic acyl transferase of Gentiana triflora var. japonica.
 XX
 KM Aromatic acyl transferase; transformation; anthocyanin pigment;
 KM plants; acylation; colour; tone; colouration; colour change;
 KM Gentiana triflora; Petunia hybrida; Perilla octinoides;
 KM Senecio cruentus; Lavandula angustifolia.
 XX
 OS Gentiana triflora var. japonica (clone pcart106).
 XX
 PN W09625500-A1.
 XX
 PD 22-AUG-1996.
 XX
 PF 16-FEB-1996; 96WO-JP00348.
 XX
 PR 30-JAN-1996; 96JP-0046534.
 PR 17-FEB-1995; 95JP-0067159.
 PR 29-JUN-1995; 95JP-0196915.
 XX
 PA (SUNR) SUNTORY LTD.
 XX
 PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
 PI Nakao M, Tanaka Y, Yonekura K;
 XX
 DR WPI: 1996-393401/39.
 DR N-PSDB: T37309.
 XX
 PT DNA coding for aromatic acyl transferase - for transforming plants
 PT which produce anthocyanin pigments and thus altering colour tone,
 PT e.g. of flowers
 XX
 PS Claim 4; Page 57-61; 94pp; Japanese.
 XX
 CC Vectors containing DNA fragments encoding proteins of plant origin
 CC with aromatic acyl transferase activity may be used to transform
 CC plants which produce anthocyanin pigments. The aromatic acyl
 CC transferase acylates the pigments in the flower resulting in colour
 CC tone changes and allowing new colourations to be produced. Six
 CC specific DNA sequences encoding aromatic acyl transferase from
 CC different plants are described in T37308-T37313.
 XX
 SQ Sequence 479 AA;

Query Match 100.0%; Score 40; DB 17; Length 479;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
 DB 415 dfgwgk 420

RESULT 11

Y51235 standard; peptide: 10 AA.

Y51235;

07-APR-2000 (first entry)

C. roseus acyl-transferase conserved peptide motif #2.

Acyltransferase; plant; RACE: screening; anti-DAT antibody;
 deacetylindoline 4-O-acetyltransferase.

Catharanthus roseus.

CA2231621-A1.

10-SEP-1999.

10-MAR-1998; 98CA-2231621.

10-MAR-1998; 98CA-2231621.

(UWMO-) UNIV MONTREAL.

St-Pierre B, De Luca V;

WPI; 2000-087682/08.

Novel amino acid consensus sequence useful for the identification of
 acyltransferase enzymes in plants -

Claim 5; Page 4; 57pp; English.

This invention describes a novel consensus sequence (1) for the
 identification of acyltransferase enzyme in plants. (1) comprises
 His-X₁-X₂-X₃-Asp-X₄ where X₁ - an amino acid comprising Ala, Arg,
 Thr, Cys, Asn, His, Met, Ile, Leu, Val or Lys; X₂ - an amino acid
 comprising Val, Ile, Leu or Met; X₃ - an amino acid comprising Cys,
 Phe, Val, Leu, Met, Ala, Ile, Met or Thr; and X₄ - an amino acid
 comprising Gly, Ala, Ile, Met or Val. The invention also describes a
 novel method for the identification of acyltransferase enzyme in plants
 comprising: (a) the steps of (1) DNA amplification using PCR primers
 specific for (1) coupled to 3'- or 5'- RACE (rapid amplification of
 cDNA ends) protocols; (11) direct screening of cDNA and genomic
 libraries using degenerate primers based on (1); (111) screening of
 cDNA expression libraries using antibodies specific for (1); or (1v)
 screening of cDNA expression libraries using anti-DAT (deacetylindoline
 4-O-acetyltransferase) antibodies; (b) identifying in clones of (1);
 (c) identifying homology to the DAT gene family; and (d) identifying an
 open reading frame encoding a protein with a molecular weight between 45
 and 54kDa. This sequence represents a plant acyl-transferase conserved
 motif which is described in the method of the invention.

Sequence 10 AA:

Query Match 92.5%; Score 37; DB 21; Length 10;

Best Local Similarity 83.3%; Pred. No. 1.8;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6

|||||

DB 1 dfgwgk 6

RESULT 12

Y68644

ID Y68644 standard; Protein: 391 AA.

XX Y68644;

05-MAY-2000 (first entry)

Amino acid sequence of the fiber protein of Adenovirus serotype 13.

Chimaeric adenovirus; gene therapy; antigenicity; fiber protein;

serotype 13; penton protein; hexon protein.

Adenovirus.

Key Location/Qualifiers

FT MISC-difference 1..31 /note= "part of the tail of adenovirus serotype 5"

FT MISC-difference 1 /note= "not specified"

FT MISC-difference 2 /note= "not specified"

FT MISC-difference 3 /note= "not specified"

FT MISC-difference 4 /note= "not specified"

FT MISC-difference 5 /note= "not specified"

FT MISC-difference 23 /note= "not specified"

FT MISC-difference 41 /note= "not specified"

FT MISC-difference 43 /note= "not specified"

FT MISC-difference 49 /note= "not specified"

FT MISC-difference 385 /note= "not specified"

WO200003029-A2.

20-JAN-2000.

08-JUL-1999; 99WO-NL00436.

08-JUL-1998; 98EP-0202297.

(INTR-) INTROGENE BV.

Havenga M, Vogels R, Bout A;

WPI; 2000-171149/15.

New chimaeric adenoviruses containing a genome derived from different
 adenovirus serotypes, useful in gene therapy -

Example 2; Fig 7; 92pp; English.

Y68642-70 represent the amino acid sequences of the fiber proteins
 of Adenovirus serotypes 8, 9, 13, 14, 20, 23, 24, 25, 27, 28, 29, 30,
 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 49 and 51.

The proteins are used in the course of the invention to construct
 chimaeric adenoviruses with reduced antigenicity. The chimaeric
 adenoviruses comprise at least part of a fiber protein of an adenovirus
 serotype providing the chimaeric virus with a desired host range and at
 least part of a penton or hexon protein from another, less antigenic,
 serotype. The chimaeric adenoviruses are useful for gene therapy,

especially where repeated delivery is required. Adenoviruses of the
 invention are useful can be constructed to have a desired host range and
 a diminished capability to raise neutralizing antibodies, an absence of,
 or decreased infection of, antigen presenting cells of the immune system
 (e.g. macrophages), and an ability to escape trapping in the liver
 through increased target cell specificity.

Sequence 391 AA:

XX

XX

XX

Query Match 87.5%; Score 35; DB 21; Length 391;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGMK 6
 :|||||
 Db 365 nfgwgk 370

RESULT 13

Y68655
 ID Y68655 standard; Protein: 391 AA.

AC Y68655;

DT 05-MAY-2000 (first entry)

DE Amino acid sequence of the fiber protein of Adenovirus serotype 33.

KW Chimeric adenovirus; gene therapy; antigenicity; fiber protein;

KM serotype 33; penton protein; hexon protein.

OS Adenovirus.

FT Key Location/Qualifiers

FT MISC-difference 1..31

FT /note- "part of the tail of adenovirus serotype 5"

PN W0200003029-A2.

PD 20-JAN-2000.

PF 08-JUL-1999; 99MO-NL00436.

PR 08-JUL-1998; 98EP-0202297.

PA (INTR-) INTROGENE BV.

PI Havenga M, Vogels R, Bout A;

DR WPI: 2000-171149/15.

PT New chimeric adenoviruses containing a genome derived from different
 PT adenovirus serotypes, useful in gene therapy -

PS Example 2; Fig 7; 92pp; English.

CC Y68642-70 represent the amino acid sequences of the fiber proteins
 CC of Adenovirus serotypes 8, 9, 13, 14, 20, 23, 24, 25, 27, 28, 29, 30,
 CC 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 49 and 51.
 CC The proteins are used in the course of the invention to construct
 CC chimeric adenoviruses with reduced antigenicity. The chimeric
 CC adenoviruses comprise at least part of a fiber protein of an adenovirus
 CC serotype providing the chimeric virus with a desired host range and at
 CC least part of a penton or hexon protein from another, less antigenic,
 CC serotype. The chimeric adenoviruses are useful for gene therapy,
 CC especially where repeated delivery is required. Adenoviruses of the
 CC invention are useful can be constructed to have a desired host range and
 CC a diminished capability to raise neutralizing antibodies, an absence of,
 CC or decreased infection of, antigen presenting cells of the immune system
 CC (e.g. macrophages), and an ability to escape trapping in the liver
 CC through increased target cell specificity.

XX Sequence 391 AA;

SO

Query Match 87.5%; Score 35; DB 21; Length 391;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGMK 6

Db 365 nfgwgk 370
 :|||||

RESULT 14

Y17400
 ID Y17400 standard; Protein: 433 AA.

AC Y17400;

DT 22-JUL-1999 (first entry)

DE Clarkia breweri benzylalcohol acetyl transferase.

KW Clarkia breweri; Clarkia conclama; benzylalcohol acetyl transferase;

KM BEAT; biosynthesis; acetyl transfer; acetyl CoA; benzylacetate; scent;

OS Clarkia breweri.

PN W09923226-A1.

PD 14-MAY-1999.

PF 28-OCT-1998; 98MO-US23023.

PR 30-OCT-1997; 97US-0063579.

PA (UNMI) UNIV MICHIGAN.

PI Dudareva N, Pichersky E;

DR WPI: 1999-326989/27.

DR N-PSDB: X56425.

PT New isolated benzylalcohol acetyl transferase

PS Claim 10; Page 146-147; 161pp; English.

CC The present sequence represents benzylalcohol acetyl transferase (BEAT)
 CC from Clarkia breweri plants. BEAT products can be used for producing
 CC plants with altered flavour and fragrance characteristics. BEAT nucleic
 CC acids can be used for increasing the synthesis of benzylacetate in a
 CC plant or for decreasing the benzylalcohol acetate content of plant
 CC cells.

SO Sequence 433 AA;

Query Match 87.5%; Score 35; DB 20; Length 433;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGMK 5
 :|||||
 Db 377 dfwgk 381

RESULT 15

Y17401
 ID Y17401 standard; Protein: 438 AA.

AC Y17401;

DT 22-JUL-1999 (first entry)

DE Clarkia conclama benzylalcohol acetyl transferase 1.

KW Clarkia breweri; Clarkia conclama; benzylalcohol acetyl transferase;

KM BEAT; biosynthesis; acetyl transfer; acetyl CoA; benzylacetate; scent;

OS Clarkia conclama.

XX
PN WO9923226-A1.
XX
PD 14-MAY-1999.
XX
PF 28-OCT-1998; 98WO-US23023.
XX
PR 30-OCT-1997; 97US-0063579.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Dudareva N, Pichersky E;
XX
DR WPI; 1999-326989/27.
DR N-PSDB; X56426.
XX
PT New isolated benzylalcohol acetyl transferase
XX
PS Claim 11: Page 148-150; 161pp; English.
XX
CC The present sequence is a benzylalcohol acetyl transferase (BEAT)
CC related protein from *Clarkia concinna* plants. BEAT products can be used
CC for producing plants with altered flavour and fragrance characteristics.
CC BEAT nucleic acids can be used for increasing the synthesis of
CC benzylacetate in a plant or for decreasing the benzylalcohol acetate
CC content of plant cells.
XX
SQ Sequence 438 AA;

Query Match 87.5%; Score 35; DB 20; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DFGMG 5
|||
Db 374 dfwg 378

Search completed: October 20, 2000, 01:35:34
Job time: 142 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 03:16:45 ; Search time 151.87 Seconds
(without alignments)
42.051 Million cell updates/sec

Title: us-08-894-356c-22

Perfect score: 17

Sequence: 1 GAYTYGNTGGGNAA 17

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14.2	83.5	17	T37314	Degenerate primer
2	14.2	83.5	23	AL3989	Geranylgeranyl dip
3	14.2	83.5	1080	Q11802	Clone PTOM36. Lyc
4	14.2	83.5	1080	Q46682	PTOM36 CDNA clone.
5	14.2	83.5	1096	V07891	Nucleotide sequenc
6	14.2	83.5	1479	T37311	Aromatic acyl tran
7	14.2	83.5	1508	T37312	Aromatic acyl tran
8	14.2	83.5	1518	T37313	Melon ripening-rel
9	14.2	83.5	1526	T89415	Anther specific cd
10	14.2	83.5	1542	Q54685	Aromatic acyl tran
11	14.2	83.5	1605	T37310	Aromatic acyl tran
12	14.2	83.5	1622	T37309	Aromatic acyl tran

13	14.2	83.5	1703	17	T37308	Aromatic acyl tran
14	14.2	83.5	3706	15	Q54693	Anther specific ge
15	14.2	83.5	3706	15	Q58340	Tobacco Ant32 gene
16	13.2	77.6	300	20	T23671	Human gene express
17	13.2	77.6	311	16	T23671	Human gene signatu
18	13.2	77.6	652	21	Z80421	Human colon cancer
19	13.2	77.6	680	18	X30844	Streptococcus pneu
20	13.2	77.6	742	21	A01753	Human colon cancer
21	13.2	77.6	1155	21	Z53471	Neisseria meningit
22	13.2	77.6	1449	20	Z12045	Neisseria meningit
23	13.2	77.6	1564	20	X58425	Clarkia breweri be
24	13.2	77.6	1831	20	X58427	Clarkia concinna b
25	13.2	77.6	3007	20	X58428	Clarkia concinna b
26	13.2	77.6	3036	20	X58426	Clarkia concinna b
27	13.2	77.6	28473	19	V52216	Streptococcus pneu
28	13	76.5	356	14	Q60522	Human brain expres
29	13	76.5	574	19	V13080	Cellulase activity
30	13	76.5	4089	17	T16857	Bovine endothelial
31	13	76.5	4097	16	Q94235	Endothelial nitrog
32	12.6	74.1	211	19	V15061	Xylanase activity
33	12.6	74.1	212	19	V15054	Xylanase activity
34	12.6	74.1	274	21	A31059	Plant microsatelli
35	12.6	74.1	311	20	V90712	Nucleotide sequenc
36	12.6	74.1	404	20	X41193	Human secreted pro
37	12.6	74.1	459	20	V86544	EST clone A2285
38	12.6	74.1	462	19	X14485	H. pylori GHPO 111
39	12.6	74.1	508	21	A29744	Human gamma 3 heav
40	12.6	74.1	508	21	Z39882	Human gamma 1 heav
41	12.6	74.1	543	20	V90626	Nucleotide sequenc
42	12.6	74.1	709	19	V31289	E. coli J96 pathog
43	12.6	74.1	747	20	Z16282	Human gene express
44	12.6	74.1	747	20	Z16088	Human gene express
45	12.6	74.1	839	20	V90681	Nucleotide sequenc

ALIGNMENTS

RESULT 1	
T37314	
ID T37314 standard; DNA; 17 bp.	
XX	
AC T37314;	
XX	
DT 06-FEB-1997 (first entry)	
XX	
DE Degenerate primer for identifying aromatic acyl transferase sequence.	
XX	
KW Aromatic acyl transferase; transformation; anthocyanin pigment;	
KW plants; acylation; colour; tone; colouration; colour change;	
KW Gentiana triflora; Petunia hybrida; Perilla oclimoides;	
KW Scenecio cruentus; Lavandula angustifolia; ss.	
XX	
OS Synthetic.	
XX	
FH Key	Location/Qualifiers
FT misc_feature	9
FT	/*tag= a
FT	/mod_base= I
FT	15
FT	/*tag= b
FT	/mod_base= I
XX	
WO9625500-A1.	
XX	
PD 22-AUG-1996.	
XX	
PF 16-FEB-1996;	96WO-JP00348.
XX	
PR 30-JAN-1996;	96JP-0046534.
PR 17-FEB-1995;	95JP-0067159.
PR 29-JUN-1995;	95JP-0196915.
XX	


```

Db      866 gatttggatgggggaaa 882
      ||:||||| ||||| ||
RESULT  4
ID      Q46682
XX      Q46682 standard; cDNA; 1080 BP.
AC      Q46682;
XX
DT      18-JAN-1994 (first entry)
DE      PTOM36 cDNA clone.
XX
KW      Transgenic fruit production; prodn.; increased solids content;
KW      higher reducing sugar content; tomato; tomatoes; plant cultivation;
KW      improved; paste; soup; ss.
XX
OS      Synthetic.
XX
PN      WO9314212-A.
XX
PD      22-JUL-1993.
XX
PF      08-JAN-1993; 93WO-GB000021.
XX
PR      10-JAN-1992; 92GB-0000520.
XX
PA      (ZENE ) ZENECA LTD.
XX
PI      Bird CR, Boniwell JM;
XX
WPI:    1993-243227/30.
XX
PT      Transgenic fruit with increased solids content - is obt'd. by
PT      cultivating plants, esp. tomatoes, in which expression of genes
PT      homologous to PTOM36 is inhibited
XX
PS      Disclosure; Fig 1: 26pp; English.
XX
CC      The sequence is that of the cDNA clone PTOM36 which is believed to
CC      encode a cytoplasmic protein of approximately 52000 daltons
CC      involved in the ripening of tomatoes. The clone is 1069 bases long
CC      with an open reading frame of 271 codons. DNA homologous to PTOM36
CC      may be used in a construct that inhibits expression of genes
CC      homologous to PTOM36 during ripening. Fruit from plants transformed
CC      with this construct have increased solids content and a higher
CC      content of reducing sugars, e.g. glucose and fructose. The fruit
CC      may be mangoes, peaches, apples, pears, strawberries, bananas,
CC      melons and esp. tomatoes, which are useful in the prodn. of
CC      improved tomato paste and soup.
XX
SQ      Sequence 1080 BP; 356 A; 141 C; 208 G; 375 T; 0 other;

Query Match      83.5%; Score 14.2; DB 14; Length 1080;
Best Local Similarity 76.5%; Pred. No. 44;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy      1 GAYTTYGGNTGGGNAA 17
      ||:||||| ||||| ||
Db      866 gatttggatgggggaaa 882

RESULT  5
V07891
XX      V07891 standard; cDNA; 1096 BP.
AC      V07891;
XX
DT      29-JAN-1999 (first entry)
DE      Nucleotide sequence of the fiber protein encoded by adenovirus Ad-36p.

```

```

XX      Adenovirus; Ad-36p fiber protein; viral obesity; vaccine; ss.
KW      Mastadenovirus.
OS
FH      Key
FT      CDS
FT      Location/Qualifiers
FT      1..1096
FT      /*tag= a
FT      /product= "Ad-36p fiber protein"
XX
PN      WO9844946-A1.
XX
PD      15-OCT-1998.
XX
PF      06-APR-1998; 98WO-US06730.
XX
PR      04-APR-1997; 97US-0042942.
XX
PA      (OBET-) OBETECH LLC.
XX
PI      Atkinson RL, Dhurandhar NV;
XX
WPI:    1998-568305/48.
XX
PT      Determining if obesity in a person is caused by Ad-36 virus - and
PT      providing the basis for treatment or prevention of obesity-causing,
PT      cholesterol reducing adenovirus, using the purified variant, Ad-36p
XX
PS      Claim 1: Pages 18-19; 24pp; English.
XX
CC      This is the nucleotide sequence encoding the adenovirus Ad-36p fiber
CC      protein used in the method of the invention to determine if a person
CC      is suffering viral obesity. The method is used to determine whether
CC      obesity in a person has a viral basis. Ad-36p can be used as a basis
CC      of a vaccine to prevent viral-based obesity.
XX
SQ      Sequence 1096 BP; 353 A; 215 C; 227 G; 301 T; 0 other;

Query Match      83.5%; Score 14.2; DB 19; Length 1096;
Best Local Similarity 76.5%; Pred. No. 44;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy      1 GAYTTYGGNTGGGNAA 17
      ||:||||| ||||| ||
Db      1013 gatttggatgggggtaa 1029

RESULT  6
T37311
ID      T37311 standard; cDNA to mRNA; 1479 BP.
XX
AC      T37311;
XX
DT      06-FEB-1997 (first entry)
DE      Aromatic acyl transferase coding sequence.
XX
KW      Aromatic acyl transferase; transformation; anthocyanin pigment;
KW      plants; acylation; colour; tone; colouration; colour change;
KW      Gentiana triflora; Petunia hybrida; Perilla ocimoides;
KW      Senecio cruentus; Lavandula angustifolia; ds.
XX
OS      Perilla ocimoides (Clone pSAT208).
XX
FH      Key
FT      CDS
FT      Location/Qualifiers
FT      3..1343
FT      /*tag= a
FT      /product= Aromatic acyl transferase.
XX
PN      WO9625500-A1.
XX
PD      22-AUG-1996.

```

```

XX PF 16-FEB-1996; 96WO-JP00348.
XX PR 30-JAN-1996; 96JP-0046534.
XX PR 17-FEB-1995; 95JP-0067159.
XX PR 29-JUN-1995; 95JP-0196915.
XX PA (SUNR ) SUNTORY LTD.
XX PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
XX PI Nakao M, Tanaka Y, Yonekura K;
XX DR WPI; 1996-393401/39.
XX DR P-PSDB; W04725.
XX XX
XX PT DNA coding for aromatic acyl transferase - for transforming plants
XX PT which produce anthocyanin pigments and thus altering colour tone,
XX PT e.g. of flowers
XX PS Claim 4; Page 65-69; 94pp; Japanese.
XX CC Vectors containing DNA fragments encoding proteins of plant origin
XX CC with aromatic acyl transferase activity may be used to transform
XX CC plants which produce anthocyanin pigments. The aromatic acyl
XX CC transferase acylates the pigments in the flower resulting in colour
XX CC tone changes and allowing new colourations to be produced. Six
XX CC specific DNA sequences encoding aromatic acyl transferase from
XX CC different plants are described in T37308-T37313.
XX SQ Sequence 1479 BP; 420 A; 316 C; 331 G; 411 T; 1 other;

Query Match 83.5%; Score 14.2; DB 17; Length 1479;
Best Local Similarity 76.5%; Pred. No. 46;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GAYTTYGGNTGGGNAA 17
   ||:|||| ||||| ||
Db 1167 gatttggatggggaa 1183

RESULT 7
T37312
ID T37312 standard; cDNA to mRNA; 1508 BP.
XX AC
XX AC T37312;
XX DT 06-FEB-1997 (first entry)
XX DE Aromatic acyl transferase coding sequence.
XX KW Aromatic acyl transferase; transformation; anthocyanin pigment;
XX KW plants; acylation; colour; tone; colouration; colour change;
XX KW Gentiana triflora; Petunia hybrida; Perilla oclmoides;
XX KW Senecio cruentus; Lavandula angustifolia; ds.
XX OS Senecio cruentus (Clone pCAT8).
XX FH Key Location/Qualifiers
XX FT CDS 3..1367
XX FT /*tag= a
XX FT /product= Aromatic acyl transferase.
XX PN WO9625500-A1.
XX PD 22-AUG-1996.
XX PF 16-FEB-1996; 96WO-JP00348.
XX PR 30-JAN-1996; 96JP-0046534.
XX PR 17-FEB-1995; 95JP-0067159.
XX PR 29-JUN-1995; 95JP-0196915.
XX PA (SUNR ) SUNTORY LTD.
XX PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
XX PI Nakao M, Tanaka Y, Yonekura K;
XX DR WPI; 1996-393401/39.
XX DR P-PSDB; W04725.
XX XX

```

```

PA (SUNR ) SUNTORY LTD.
XX Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
XX PI Nakao M, Tanaka Y, Yonekura K;
XX DR WPI; 1996-393401/39.
XX DR P-PSDB; W04726.
XX XX
XX PT DNA coding for aromatic acyl transferase - for transforming plants
XX PT which produce anthocyanin pigments and thus altering colour tone,
XX PT e.g. of flowers
XX PS Claim 4; Page 69-72; 94pp; Japanese.
XX CC Vectors containing DNA fragments encoding proteins of plant origin
XX CC with aromatic acyl transferase activity may be used to transform
XX CC plants which produce anthocyanin pigments. The aromatic acyl
XX CC transferase acylates the pigments in the flower resulting in colour
XX CC tone changes and allowing new colourations to be produced. Six
XX CC specific DNA sequences encoding aromatic acyl transferase from
XX CC different plants are described in T37308-T37313.
XX SQ Sequence 1508 BP; 442 A; 294 C; 295 G; 477 T; 0 other;

Query Match 83.5%; Score 14.2; DB 17; Length 1508;
Best Local Similarity 76.5%; Pred. No. 46;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GAYTTYGGNTGGGNAA 17
   ||:|||| ||||| ||
Db 1182 gatttggatggggaa 1198

RESULT 8
T37313
ID T37313 standard; cDNA to mRNA; 1518 BP.
XX AC
XX AC T37313;
XX DT 06-FEB-1997 (first entry)
XX DE Aromatic acyl transferase coding sequence.
XX KW Aromatic acyl transferase; transformation; anthocyanin pigment;
XX KW plants; acylation; colour; tone; colouration; colour change;
XX KW Gentiana triflora; Petunia hybrida; Perilla oclmoides;
XX KW Senecio cruentus; Lavandula angustifolia; ds.
XX OS Lavandula angustifolia (Clone pLAT21).
XX FH Key Location/Qualifiers
XX FT CDS 1..1355
XX FT /*tag= a
XX FT /product= Aromatic acyl transferase.
XX PN WO9625500-A1.
XX PD 22-AUG-1996.
XX PF 16-FEB-1996; 96WO-JP00348.
XX PR 30-JAN-1996; 96JP-0046534.
XX PR 17-FEB-1995; 95JP-0067159.
XX PR 29-JUN-1995; 95JP-0196915.
XX PA (SUNR ) SUNTORY LTD.
XX PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
XX PI Nakao M, Tanaka Y, Yonekura K;
XX DR WPI; 1996-393401/39.
XX XX

```

PT DNA coding for aromatic acyl transferase - for transforming plants
 PT which produce anthocyanin pigments and thus altering colour tone,
 PS e.g. of flowers
 XX
 XX Claim 4; Page 73-76; 94pp; Japanese.
 XX
 CC Vectors containing DNA fragments encoding proteins of plant origin
 CC with aromatic acyl transferase activity may be used to transform
 CC plants which produce anthocyanin pigments. The aromatic acyl
 CC transferase acylates the pigments in the flower resulting in colour
 CC tone changes and allowing new colourations to be produced. Six
 CC specific DNA sequences encoding aromatic acyl transferase from
 CC different plants are described in T37308-T37313. NOTE: This
 CC sequence is supposed to cross reference with the protein described
 CC in W04727, however there are so many discrepancies between the
 CC polypeptide deduced from this sequence and the polypeptide given in
 CC the specification and described in W04727 that the indexer decided
 CC not to cross reference the two.
 XX
 SQ Sequence 1518 BP; 384 A; 340 C; 366 G; 428 T; 0 other;

Query Match 83.5%; Score 14.2; DB 17; Length 1518;
 Best Local Similarity 76.5%; Pred. No. 46;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGGTGGGNA 17
 ||:|||||
 Db 1176 gatttggatggggta 1192

RESULT 9
 T89415
 ID T89415 standard; cDNA; 1526 BP.
 AC T89415;
 XX
 DT 15-APR-1998 (first entry)
 XX
 DE Melon ripening-related cDNA clone MEL2.
 XX
 KW Melon cultivar Canteloupe charantais; fruit ripening control;
 KW ethylene; MEL2 promoter; ss.
 XX
 OS Cucumis melo.
 XX
 FH Key Location/Qualifiers
 FT 3'UTR 1385..1526
 FT /*tag= a
 FT /note= "Contains a putative polyadenylation signal"
 PT
 PN W09737023-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 24-MAR-1997; 97WO-GB00824.
 XX
 PR 02-APR-1996; 96GB-0006906.
 XX
 PA (ZENE) ZENECA LTD.
 XX
 PI Aggelis A, Greiferson D, John I, Karvouni Z;
 XX WPI: 1997-503108/46.
 DR
 XX
 CC DNA related with fruit ripening - can be used to genetically modify
 PT especially melons to control ripening
 XX
 PS Claim 1; Pages 18-19; 38pp; English.
 XX
 CC This cDNA sequence is from the MEL2 clone which produces a
 CC novel ripening-related product from Cucumis melo. It is one of two
 CC specified sequences (MEL2) 1526 bp or (MEL7) 686 bp which were

CC isolated from a melon ripe fruit cDNA library. MEL2 is not a
 CC full-length clone; it lacks the initiation codon for the amino terminus.
 CC The cDNA allows control of the ripening of fruit, especially melons.
 XX
 SQ Sequence 1526 BP; 439 A; 297 C; 331 G; 448 T; 2 U; 9 other;

Query Match 83.5%; Score 14.2; DB 18; Length 1526;
 Best Local Similarity 76.5%; Pred. No. 46;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGGTGGGNA 17
 ||:|||||
 Db 1136 gacttggatggggaa 1152

RESULT 10
 Q54685
 ID Q54685 standard; cDNA; 1542 BP.
 XX
 AC Q54685;
 XX
 DT 06-JUL-1994 (first entry)
 XX
 DE Anther specific cDNA clone ant32.
 XX
 KW Transgenic plants; male sterility; pollen; sterile;
 KW self-pollination; seed; hybrid; toxin-A; peIE; T-urf13; Gln; laal;
 KW CyTA; toxin; Nicotiana tabacum; Diptheria; Erwinia chrysanthemi;
 KW phage Mu; Psuedomonas syringae; Bacillus thuringiensis; anther;
 KW breeding; ss.
 XX
 OS Nicotiana tabacum.
 XX
 FH Key Location/Qualifiers
 FT CDS 66..1412
 FT /*tag= a
 XX
 PN EP578611-A.
 XX
 PD 12-JAN-1994.
 XX
 PF 24-JUN-1993; 93EP-0810455.
 XX
 PR 02-JUL-1992; 92US-0908242.
 XX
 PA (CIBA) CIBA GEIGY AG.
 XX
 PT Crossland LD, Tuttle AB;
 PT WPI: 1994-010428/02.
 DR P-PSDB; R47475.
 DR
 XX
 CC Anther-specific cDNA, genomic and recombinant DNA - produce
 PT transgenic male-sterile plants, which prevents self-pollination,
 PT in hybrid seed prodn.
 XX
 PS Claim 1; Page 21-24; 75pp; English.
 XX
 CC Anther specific cDNA or genomic sequences can be used to identify
 CC and isolate anther specific promoters. The anther specific promoter
 CC can then be cloned into a recombinant construct and used to express
 CC heterologous genes. Preferred heterologous genes include Diptheria
 CC toxin A-chain gene; peptate lyase gene peIE from Erwinia
 CC chrysanthemi; T-urf13 from cms-T maize mitochondrial genomes; the
 CC Gln recombinase gene from phage Mu; the indole acetic acid-lyase
 CC synthetase gene from Pseudomonas syringae and the CyTA toxin gene
 CC from Bacillus thuringiensis israeliensis. All of these genes when
 CC expressed in anther tissue will result in the inability of the
 CC plant to produce viable pollen. Transformation of plants with such
 CC a recombinant construct can produce transgenic, male sterile plants.
 CC Male sterility is important in the production of hybrid seeds as it
 CC prevents self pollination which hinders breeding and hybrid seed

CC production.

XX Sequence 1542 BP; 473 A; 293 C; 332 G; 444 T; 0 other;

SQ Query Match 83.5%; Score 14.2; DB 15; Length 1542;
Best Local Similarity 76.5%; Pred. No. 46;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAYTTYGGNTGGGNAA 17

Db 1239 gatttggatggggaaa 1255

RESULT 11

T37310
ID T37310 standard; cDNA to mRNA; 1605 BP.

XX AC T37310;

XX DT 06-FEB-1997 (first entry)

XX DE Aromatic acyl transferase coding sequence.

XX KW Aromatic acyl transferase; transformation; anthocyanin pigment;
XX plants; acylation; colour; tone; colouration; colour change;
XX Gentiana triflora; Petunia hybrida; Perilla ocimoides;
XX Scenecio cruentus; Lavandula angustifolia; ds.

XX OS Petunia hybrida (Clone pPAT48).

XX FH Key Location/Qualifiers
XX CDS 67..1413
FT /*tag= a
FT /product= Aromatic acyl transferase.

XX PN WO9625500-A1.

XX PD 22-AUG-1996.

XX PF 16-FEB-1996; 96WO-JP00348.

XX PR 30-JAN-1996; 96JP-0046534.

XX PR 17-FEB-1995; 95JP-0067159.

XX PR 29-JUN-1995; 95JP-0196915.

XX PA (SUNR) SUNTORY LTD.

XX PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;

XX PI Nakao M, Tanaka Y, Yonekura K;

XX WPI: 1996-393401/39.

XX DR P-PSDB; W04724.

XX PT DNA coding for aromatic acyl transferase - for transforming plants
XX which produce anthocyanin pigments and thus altering colour tone,
XX e.g. of flowers

XX PS Claim 4; Page 61-65; 94pp; Japanese.

XX CC Vectors containing DNA fragments encoding proteins of plant origin
XX with aromatic acyl transferase activity may be used to transform
XX plants which produce anthocyanin pigments. The aromatic acyl
XX transferase acylates the pigments in the flower resulting in colour
XX tone changes and allowing new colourations to be produced. Six
XX specific DNA sequences encoding aromatic acyl transferase from
XX different plants are described in T37308-T37313.

XX SQ Sequence 1605 BP; 477 A; 319 C; 375 G; 434 T; 0 other;

Query Match

Best Local Similarity 83.5%; Score 14.2; DB 17; Length 1605;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAYTTYGGNTGGGNAA 17

Db 1234 gatttggatggggaaa 1250

RESULT 12

T37309
ID T37309 standard; cDNA to mRNA; 1622 BP.

XX AC T37309;

XX DT 06-FEB-1997 (first entry)

XX DE Aromatic acyl transferase coding sequence.

XX KW Aromatic acyl transferase; transformation; anthocyanin pigment;
XX plants; acylation; colour; tone; colouration; colour change;
XX Gentiana triflora; Petunia hybrida; Perilla ocimoides;
XX Scenecio cruentus; Lavandula angustifolia; ds.

XX OS Gentiana triflora var. japonica (Clone pGAT106).

XX FH Key Location/Qualifiers
XX CDS 6..1415
FT /*tag= a
FT /product= Aromatic acyl transferase.

XX FT misc_feature 701..703

XX FT /*tag= b

XX FT /*transl_except= ATC encodes Leu

XX FT misc_feature 1322..1324

XX FT /*tag= c

XX FT /*transl_except= AAT encodes Asp

XX PN WO9625500-A1.

XX PD 22-AUG-1996.

XX PF 16-FEB-1996; 96WO-JP00348.

XX PR 30-JAN-1996; 96JP-0046534.

XX PR 17-FEB-1995; 95JP-0067159.

XX PR 29-JUN-1995; 95JP-0196915.

XX PA (SUNR) SUNTORY LTD.

XX PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;

XX PI Nakao M, Tanaka Y, Yonekura K;

XX WPI: 1996-393401/39.

XX DR P-PSDB; W04723.

XX PT DNA coding for aromatic acyl transferase - for transforming plants

XX which produce anthocyanin pigments and thus altering colour tone,

XX e.g. of flowers

XX PS Claim 4; Page 57-61; 94pp; Japanese.

XX CC Vectors containing DNA fragments encoding proteins of plant origin

XX with aromatic acyl transferase activity may be used to transform

XX plants which produce anthocyanin pigments. The aromatic acyl

XX transferase acylates the pigments in the flower resulting in colour

XX tone changes and allowing new colourations to be produced. Six

XX specific DNA sequences encoding aromatic acyl transferase from

XX different plants are described in T37308-T37313.

XX SQ Sequence 1622 BP; 478 A; 354 C; 313 G; 477 T; 0 other;

Query Match

Best Local Similarity 83.5%; Score 14.2; DB 17; Length 1622;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Job time: 5994 sec

```
RESULT 15
Q58340
ID Q58340 standard; DNA; 3706 BP.
XX
AC Q58340;
XX
DT 27-OCT-1994 (first entry)
XX
DE Tobacco Ant32 genomic clone sequence.
XX
KW ant32 promoter; genomic clone; male-sterile plants;
KW fusion genes; ss.
XX
OS Nicotiana tabacum.
XX
FH Key Location/Qualifiers
FT TATA_signal 1971..1975
FT FT /*tag= a
FT misc_feature 2009
FT FT /*tag= b
FT FT /note= "putative transcription start site"
FT 5'UTR 1..2075
FT FT /*tag= c
FT FT /note= "ant32 promoter can be obtained from
FT the 2.0kb 5' flanking region of the
FT ant32 clone"
FT CDS 2076..3422
FT FT /*tag= d
XX
EP589841-A.
XX
30-MAR-1994.
XX
15-SEP-1993; 93EP-0810654.
XX
24-SEP-1992; 92US-0950348.
XX
(CIBA ) CIBA GEIGY AG.
XX
Crossland LD, Stein JI, Tuttle A;
XX
WPI; 1994-119827/15.
XX
P-PSDB; R46929.
XX
Dual method for producing male-sterile plants - comprises
PT crossing 2 genetically transformed plants, useful for producing
PT hybrid seed
XX
Example 8; Page 22-28; 30pp; English.
XX
Q58340 is an anther specific clone from Nicotiana tabacum.
CC The ant32 promoter can be operably linked to, eg. T7 RNA
CC polymerase gene or to GAL4/VPI fusion gene in order to
CC produce male-sterile plants. The inventors claim that
CC male-sterile plants finally produced can be used to increase
CC hybrid vigour in plants, e.g. maize.
XX
SQ Sequence 3706 BP; 1097 A; 653 C; 709 G; 1247 T; 0 other;
```

```
Query Match 83.5%; Score 14.2; DB 15; Length 3706;
Best Local Similarity 76.5%; Pred. No. 50;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAYTTYGGNTGGGNAA 17
   ||:|||||
Db 3249 gattttggtgggaaaa 3265
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 03:12:31 ; Search time 103.47 Seconds
(Without alignments)
24.849 Million cell updates/sec

Title: US-08-894-356c-22

Sequence: 1 GAYTTGGTGGGNA 17

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	14.2	83.5	23	US-09-187-050-6	Sequence 6, Appl1
2	14.2	83.5	1080	US-07-598-873-1	Sequence 1, Appl1
3	14.2	83.5	1080	US-08-073-425-1	Sequence 1, Appl1
4	14.2	83.5	1080	US-08-396-531-1	Sequence 1, Appl1
5	14.2	83.5	1512	US-09-142-514-1	Sequence 1, Appl1
6	14.2	83.5	1542	US-08-142-514-1	Sequence 1, Appl1
7	14.2	83.5	3706	US-08-207-904-1	Sequence 16, Appl1
8	13	76.5	4089	US-07-908-245-1	Sequence 1, Appl1
9	12.6	74.1	508	US-08-822-028-74	Sequence 7, Appl1
10	12.6	74.1	912	US-08-599-171A-24	Sequence 24, Appl1
11	12.6	74.1	912	US-08-646-5908-24	Sequence 24, Appl1
12	12.6	74.1	912	US-09-069-226-24	Sequence 24, Appl1
13	12.6	74.1	930	US-08-096-182A-1	Sequence 1, Appl1
14	12.6	74.1	930	US-08-877-109-1	Sequence 1, Appl1
15	12.6	74.1	930	US-08-798-760-1	Sequence 1, Appl1
16	12.6	74.1	930	PCT-US94-08327-1	Sequence 1, Appl1
17	12.6	74.1	2344	US-07-695-472B-35	Sequence 35, Appl1
18	12.6	74.1	2347	US-08-250-740-32	Sequence 32, Appl1
19	12.6	74.1	2347	US-07-695-472B-1	Sequence 34, Appl1
20	12.6	74.1	2347	US-07-695-472B-36	Sequence 36, Appl1
21	12.6	74.1	2360	US-07-916-098A-44	Sequence 1, Appl1
22	12.6	74.1	3167	US-08-472-534-1	Sequence 1, Appl1
23	12.6	74.1	3393	US-08-295-502-1	Sequence 1, Appl1
24	12.6	74.1	3393	PCT-US95-10691-1	Sequence 1, Appl1
25	12.6	74.1	4543	US-08-519-547A-5	Sequence 5, Appl1
C 26	12.6	74.1	4543	US-08-519-547A-5	Sequence 5, Appl1

27	12.6	74.1	4741	2	US-07-695-472B-4	Sequence 4, Appl1
28	12.6	74.1	4742	5	US-08-250-740-35	Sequence 35, Appl1
C 29	12.2	71.8	531	5	US-08-784-582-59	Sequence 59, Appl1
C 30	12.2	71.8	579	1	US-07-749-446-1	Sequence 1, Appl1
C 31	12.2	71.8	603	6	PCT-US94-12873-5	Sequence 5, Appl1
32	12.2	71.8	626	5	US-08-906-769-150	Sequence 150, Appl1
C 33	12.2	71.8	782	5	US-08-906-769-150	Sequence 150, Appl1
C 34	12.2	71.8	782	3	US-07-865-878A-1	Sequence 1, Appl1
C 35	12.2	71.8	782	3	US-08-308-736A-20	Sequence 20, Appl1
C 36	12.2	71.8	782	4	US-08-603-010-1	Sequence 1, Appl1
C 37	12	70.6	96	1	US-08-120-827-89	Sequence 89, Appl1
C 38	12	70.6	96	2	US-08-478-675-89	Sequence 89, Appl1
C 39	12	70.6	226	1	US-08-090-523-30	Sequence 30, Appl1
C 40	12	70.6	226	1	US-08-398-627-30	Sequence 30, Appl1
C 41	12	70.6	4951	4	US-08-867-030B-5	Sequence 5, Appl1
C 42	12	70.6	4951	6	PCT-US95-06119-5	Sequence 5, Appl1
43	12	70.6	35100	2	US-08-306-691B-19	Sequence 19, Appl1
44	12	70.6	35100	6	PCT-US93-06251-19	Sequence 19, Appl1
C 45	11.6	68.2	565	2	US-08-592-126-141	Sequence 141, Appl1

ALIGNMENTS

RESULT 1
US-09-187-050-6/C
Sequence 6, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: MSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: non-degenerate
FEATURE:
NAME/KEY: misc difference
LOCATION: (1)..(23)
OTHER INFORMATION: No. 6043072-degenerate PCR primer
US-09-187-050-6

Query Match 83.5%; Score 14.2; DB 5; Length 23;
Best Local Similarity 76.5%; Pred. No. 5.7;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAYTTGGTGGGNA 17
Db 20 GATTTCGGCTGGGTA 4

RESULT 2
US-07-598-873-1
Sequence 1, Application US/07598873
Patent No. 5254800
GENERAL INFORMATION:
APPLICANT: BIRD, COLIN R
APPLICANT: GRIERSON, DONALD
APPLICANT: RAY, JOHN A
APPLICANT: SCHUCH, WOLFGANG W
TITLE OF INVENTION: DNA, CONSTRUCTS, CELLS AND PLANTS
TITLE OF INVENTION: DERIVED THEREFROM
NUMBER OF SEQUENCES: 5

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
;; STREET: Eleventh Floor, 1615 L Street, N.W.
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20036-5601
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/598,873
;; FILING DATE: 19901019
;; CLASSIFICATION: 800
;; ATTORNEY/AGENT INFORMATION:
;; NAME: WILSON, MARY J
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-861-3000
;; TELEFAX: 202-822-0944
;; TELEX: 6714627 CUSH
;;
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1080 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to mRNA
;; HYPOTHEICAL: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Lycopersicon esculentum
;; STRAIN: Ailsa Craig
;; DEVELOPMENTAL STAGE: Ripening
;;
US-07-598-873-1

Query Match 83.5%; Score 14.2; DB 1; Length 1080;
Best Local Similarity 76.5%; Pred. No. 10;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGGNTGGGNA 17
||:|:|:| 11111 11
DB 866 GATTTGGATGGGNA 882

RESULT 3
US-08-073-425-1
;; Sequence 1, Application US/08073425
;; Patent No. 5569829
;; GENERAL INFORMATION:
;; APPLICANT: BIRD, COLIN R
;; APPLICANT: BONIWELL, JEREMY M.
;; APPLICANT: GRIERSON, DONALD
;; APPLICANT: RAY, JOHN A
;; APPLICANT: SCHUCH, WOLFGANG W
;; TITLE OF INVENTION: PLANTS
;; NUMBER OF SEQUENCES: 5
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
;; STREET: 1100 New York Avenue, N.W.
;; CITY: WASHINGTON
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-3918
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/073,425
;; FILING DATE: 09-JUN-1993

;; CLASSIFICATION: 800
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DEAYER, DONALD B.
;; REGISTRATION NUMBER: 23,048
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-861-3000
;; TELEFAX: 202-822-0944
;; TELEX: 6714627 CUSH
;;
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1080 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to mRNA
;; HYPOTHEICAL: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Lycopersicon esculentum
;; STRAIN: Ailsa Craig
;; DEVELOPMENTAL STAGE: Ripening
;;
US-08-073-425-1

Query Match 83.5%; Score 14.2; DB 1; Length 1080;
Best Local Similarity 76.5%; Pred. No. 10;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGGNTGGGNA 17
||:|:|:| 11111 11
DB 866 GATTTGGATGGGNA 882

RESULT 4
US-08-396-531-1
;; Sequence 1, Application US/08396531
;; Patent No. 5744364
;; GENERAL INFORMATION:
;; APPLICANT: BIRD, COLIN R
;; APPLICANT: GRIERSON, DONALD
;; APPLICANT: RAY, JOHN A
;; APPLICANT: SCHUCH, WOLFGANG W
;; TITLE OF INVENTION: DNA, CONSTRUCTS, CELLS AND PLANTS
;; NUMBER OF SEQUENCES: 5
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
;; STREET: Ninth Floor, 1100 New York Avenue, N.W.
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-3918
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/396,531
;; FILING DATE:
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/847,037
;; FILING DATE: 16-APR-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KOKULIS, PAUL N.
;; REGISTRATION NUMBER: 16,773
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-861-3000
;; TELEFAX: 202-822-0944
;; TELEX: 6714627 CUSH
;;
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1080 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Lycopersicon esculentum
STRAIN: Alisa Craig
DEVELOPMENTAL STRGE: Ripening
US-08-396-531-1

Query Match 83.5%; Score 14.2; DB 2; Length 1080;
Best Local Similarity 76.5%; Pred. No. 10;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTGGNTGGGNA 17
||:|:|:|:|:|:|:|:|:|
Db 866 GATTTGGATGCGGAA 882

RESULT 5
US-09-142-514-1
Sequence 1, Application US/09142514A
Patent No. 6107548
GENERAL INFORMATION:
APPLICANT: Aggelis, Alexandros
APPLICANT: Grierson, Donald
APPLICANT: John, Isaac
APPLICANT: Karvouni, Zoi
TITLE OF INVENTION: Fruit Ripening
FILE REFERENCE: SE50111/UST
CURRENT APPLICATION NUMBER: US/09/142,514A
CURRENT FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: PCT/GB97/00824
EARLIER FILING DATE: 1997-03-24
EARLIER APPLICATION NUMBER: GB 9606906.7
EARLIER FILING DATE: 1996-04-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 1512
TYPE: DNA
ORGANISM: Cucumis melo
US-09-142-514-1

Query Match 83.5%; Score 14.2; DB 5; Length 1512;
Best Local Similarity 76.5%; Pred. No. 11;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTGGNTGGGNA 17
||:|:|:|:|:|:|:|:|:|
Db 1122 gacttgcgtgcgga 1138

RESULT 6
US-08-207-904-1
Sequence 1, Application US/08207904
Patent No. 5477002
GENERAL INFORMATION:
APPLICANT: Tuttle, AnnMarie
APPLICANT: Crossland, Lyle D.
TITLE OF INVENTION: Anthr-Specific cDNA Sequences, Genomic
TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/207,904
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/908,242
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: CGC 1624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1542 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Nicotiana tabacum
INDIVIDUAL ISOLATE: Ant32
FEATURE:
NAME/KEY: CDS
LOCATION: 66..1412
US-08-207-904-1

Query Match 83.5%; Score 14.2; DB 1; Length 1542;
Best Local Similarity 76.5%; Pred. No. 11;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTGGNTGGGNA 17
||:|:|:|:|:|:|:|:|:|
Db 1239 GATTTGGTGGGNA 1255

RESULT 7
US-08-207-904-16
Sequence 16, Application US/08207904
Patent No. 5477002
GENERAL INFORMATION:
APPLICANT: Tuttle, AnnMarie
APPLICANT: Crossland, Lyle D.
TITLE OF INVENTION: Anthr-Specific cDNA Sequences, Genomic
TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/207,904
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/908,242
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: CGC 1624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 3706 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Nicotiana tabacum
INDIVIDUAL ISOLATE: Ant32 genomic clone
IMMEDIATE SOURCE:
CLONE: PCIB950
FEATURE:
NAME/KEY: TATA_signal
LOCATION: 1971..1975
FEATURE:
NAME/KEY: CDS
LOCATION: 2076..3422
FEATURE:
NAME/KEY: misc-feature
LOCATION: 2009
OTHER INFORMATION: /note="Putative transcription
US-08-207-904-16

Query Match 83.5%; Score 14.2; DB 1; Length 3706;
Best Local Similarity 76.5%; Pred. No. 13;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGGTGGGANA 17
||:|:|:|:|:|:|:|:|:|
DB 3249 GATTTGGTGGGANA 3265

RESULT 8
US-07-908-245-1/C
Sequence 1, Application US/07908245
Patent No. 5498539
GENERAL INFORMATION:
APPLICANT: Harrison, David G.
APPLICANT: Alexander, R. Wayne
APPLICANT: Murphy, T.J.
APPLICANT: Nishida, Ken'ichi
TITLE OF INVENTION: Endothelial Nitric Oxide Synthase
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30309-4530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/908,245
FILING DATE: 19920702
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU 111
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4089 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bovine
TISSUE TYPE: Aorta
CELL TYPE: Endothelial
US-07-908-245-1

Query Match 76.5%; Score 13; DB 1; Length 4089;
Best Local Similarity 76.5%; Pred. No. 65;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGGTGGGANA 17
||:|:|:|:|:|:|:|:|:|
DB 3840 GACTTAGGTGGGANA 3824

RESULT 9
US-08-822-028-74
Sequence 74, Application US/08822028
Patent No. 5993813
GENERAL INFORMATION:
APPLICANT: MEZES, PETER S
APPLICANT: GOURLE, BRIAN B
APPLICANT: RIXON, MARK W
APPLICANT: ANDERSON, WH KERR
APPLICANT: KAPLAN, DONALD A
APPLICANT: SCHLOM, JEFFREY
TITLE OF INVENTION: A NOVEL FAMILY OF HIGH AFFINITY,
NUMBER OF SEQUENCES: 74
MODIFIED ANTIBODIES FOR CANCER TREATMENT
CORRESPONDENCE ADDRESS:
ADDRESSEE: DUANE C ULMER
STREET: P. O. BOX 1967
CITY: MIDLAND
STATE: MICHIGAN
COUNTRY: USA
ZIP: 48641-1967
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/040,687
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: ULMER, DUANE C
REGISTRATION NUMBER: 34,941
REFERENCE/DOCKET NUMBER: C-37,075C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 636-8104
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:

COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/069,226
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/599,171
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HERRON, CHARLES J.
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-38
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 912 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
US-09-069-226-24

Query Match 74.1%; Score 12.6; DB 5; Length 912;
Best Local Similarity 70.6%; Pred. No. 88;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAYTTGGNTGGGNA 17
||:|:| | | | | | | |
Db 686 GACTCGGTGGAGGAA 702

RESULT 13
US-08-096-182A-1
Sequence 1, Application US/08096182A
Patent No. 5439808
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Qi, Huijin L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hronowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
Purification and Refolding of the Outer Membrane Group B
TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,182A
FILING DATE: 23-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.

REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438,0060000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
FEATURE:
NAME/KEY: CDS
LOCATION: 1..930
US-08-096-182A-1

Query Match 74.1%; Score 12.6; DB 1; Length 930;
Best Local Similarity 70.6%; Pred. No. 88;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAYTTGGNTGGGNA 17
||:|:| | | | | | | |
Db 196 GACTCGGTGGGCA 212

RESULT 14
US-08-877-109-1
Sequence 1, Application US/08877109
Patent No. 5747287
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Qi, Huijin L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hronowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
Purification and Refolding of the Outer Membrane Group B
TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
TITLE OF INVENTION: Porin Proteins from Neisseria meningitidis
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/877,109
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/431,264
FILING DATE: 28-APR-1995
APPLICATION NUMBER: 08/096,182
FILING DATE: 23-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438,0060001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 930 base pairs
TYPE: nucleic acid

STRANDEDNESS: not relevant
TOPOLOGY: not relevant
FEATURE:
NAME/KEY: CDS
LOCATION: 1..930
US-08-877-109-1

Db 196 GACTCCGTTGGGCAA 212

Search completed: October 20, 2000, 03:12:36
Job time: 5774 sec

Query Match 74.1%; Score 12.6; DB 2; Length 930;
Best Local Similarity 70.6%; Pred. No. 88;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAYTTGNTGGGNA 17
Db 196 GACTCCGTTGGGCAA 212

RESULT 15
US-08-798-760-1
Sequence 1, Application US/08798760
Patent No. 6013267
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Qi, Huihan L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hronowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
TIME OF INVENTION: Purification and Refolding of the Outer Membrane Group B
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,760
FILING DATE: 11-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438-0060002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
FAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
FEATURE:
NAME/KEY: CDS
LOCATION: 1..930
US-08-798-760-1

Query Match 74.1%; Score 12.6; DB 5; Length 930;
Best Local Similarity 70.6%; Pred. No. 88;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAYTTGNTGGGNA 17
11:1:11111111

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 02:07:32 ; Search time 1792.79 Seconds

(without alignments)
56.628 Million cell updates/sec

Title: US-08-894-356c-22

17

Sequence: 1 GAYTYGNTGGGNA 17

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gsa1:*
89: gb_gsa2:*
90: gb_gsa3:*
91: gb_gsa4:*
92: em_gsa1:*
93: em_gsa2:*
94: em_gsa3:*
95: em_gsa4:*
96: gb_gsa5:*
97: gb_gsa6:*
98: gb_gsa7:*
99: gb_gsa8:*
100: gb_gsa9:*
101: em_gsa5:*
102: em_gsa6:*
103: em_gsa7:*
104: em_gsa8:*
105: em_gsa9:*
106: em_gsa10:*
107: em_gsa11:*
108: gb_gsa10:*
109: gb_gsa11:*
110: em_gsa12:*
111: gb_gsa12:*
112: gb_gsa13:*
113: gb_gsa14:*
114: gb_gsa15:*
115: gb_gsa16:*
116: gb_gsa17:*

117: gb_gss18:*
 118: gb_gss19:*
 119: em_gss13:*
 120: gb_gss20:*
 121: gb_gss21:*
 122: gb_gss22:*
 123: gb_gss23:*
 124: gb_gss24:*
 125: em_gss14:*
 126: em_gss15:*
 127: em_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14.2	83.5	227	11	AI494813	sa96g04.y
2	14.2	83.5	244	20	AM234169	sf22a08.y
3	14.2	83.5	247	11	AI488429	EST246768
4	14.2	83.5	249	113	A2015624	RPCI-23-2
5	14.2	83.5	258	5	AA650754	30868 Lam
6	14.2	83.5	263	5	AA651035	31149 Lam
7	14.2	83.5	278	11	AI608251	vh93f09.x
8	14.2	83.5	294	22	BB264508	BB264508
9	14.2	83.5	307	22	AM471629	s113h01.y
10	14.2	83.5	324	13	AI856623	sb39g12.y
11	14.2	83.5	333	88	AO034368	CIT-HSP-2
12	14.2	83.5	347	35	BE460581	EST412000
13	14.2	83.5	360	14	AL378020	MTB835P12
14	14.2	83.5	367	21	AM350546	GM210009A
15	14.2	83.5	389	38	R63685	Y110f04.s1
16	14.2	83.5	397	15	AI967062	496021B11
17	14.2	83.5	397	15	AU096989	AU096989
18	14.2	83.5	402	18	AV407679	GF-FV-P6E
19	14.2	83.5	405	34	BE213479	BE213479
20	14.2	83.5	409	32	BE020415	sm43h12.y
21	14.2	83.5	410	19	AV539490	AV539490
22	14.2	83.5	414	19	AM043809	wy81b04.x
23	14.2	83.5	417	12	AI748527	sb54c08.y
24	14.2	83.5	417	18	AV425743	AV425743
25	14.2	83.5	417	39	WO8160	mb42605.r1
26	14.2	83.5	424	21	AM318655	un03b01.y
27	14.2	83.5	426	35	BE461054	EST412473
28	14.2	83.5	427	3	AA394828	26611 Lam
29	14.2	83.5	428	90	AO186186	HS_3074_B
30	14.2	83.5	441	6	AA713159	AA713159
31	14.2	83.5	445	13	BE432935	EST399380
32	14.2	83.5	446	35	AI899217	EST268660
33	14.2	83.5	448	21	AM255739	ML800 pep
34	14.2	83.5	449	23	AM650512	EST328966
35	14.2	83.5	462	13	AI856260	sb39g12.x
36	14.2	83.5	469	35	BE432067	EST398596
37	14.2	83.5	473	19	AV534749	AV534749
38	14.2	83.5	478	100	AO666486	HS_5317_A
39	14.2	83.5	479	6	R08263	Y118a11.s1
40	14.2	83.5	480	6	AA754069	97GS0782
41	14.2	83.5	481	5	AA650753	30867 Lam
42	14.2	83.5	484	21	AM283831	LGL_260.C
43	14.2	83.5	484	117	AZ269810	RPCI-23-1
44	14.2	83.5	485	14	AL386583	MCBC35E10
45	14.2	83.5	486	10	AI442795	sa26f04.x

ALIGNMENTS

RESULT 1
 AI494813

LOCUS AI494813 227 bp mRNA
 DEFINITION sa96g04.y1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl004-7183 5' similar to TR-Q43583 Q43583 HSR201 PROTEIN. ;
 mRNA sequence.
 AI494813
 AI494813.1 GI:4395816
 EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 227)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelidng,J., Corryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Willson,R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Seq primer: 40RP from Gibco
 High quality sequence stop: 212.
 Location/Qualifiers
 1..227
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl004-7183"
 /clone_1lb="Gm-cl004"
 /clone_1lb="Gm-cl004"
 /tissue_type="root"
 /lab_host="X110-GOLD"
 /note="Vector: pBluescript II XR. Site.1: EcoRI; Site.2: XhoI. Root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stratagene's first-strand synthesis primer was used (GAGGAGAGAGAGAGACTGCTCCGCT)-18). After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adaptors, and phosphorylated. The xhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all xhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and xhoI, and phosphorylated. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies (9n-15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Corryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011. Phone: 520-523-1078 (Dr. Paul Keim), 520-523-7500 (Virginia H. Corryell), Fax: 520-523-7500, email: paul.keim@nau.edu, virginia.corryell@nau.edu"

BASE COUNT

52 a 39 c 68 g 68 t

ORIGIN

Query Match 83.5%; Score 14.2; EB 11; Length 227;
 Best Local Similarity 76.5%; Pred. No. 7.5e+02;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATTTGGNTGGGNA 17
 ||:|:|:|:|:|:|:|:|:|
 Db 23 GACTTGGTTGGGGAA 39

RESULT 2

AM234169 244 bp mRNA EST 17-JUL-2000
 LOCUS sf22808.y1 Gm-cl028 Glycine max cDNA CLONE GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-cl028-687 5' similar to TR:Q43583 Q43583 HSR201 PROTEIN. ;, mRNA
 sequence.
 ACCESSION AM234169
 VERSION AM234169.1 GI:6566526
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max

Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:
 Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids I;
 Fabales: Fabaceae: Papilionoideae; Glycine.
 1 (bases 1 to 244)

REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corryell,V., Rhanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Rheising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk
 R., Ritter,E., Kohn,S., Shun,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)

TITLE Public Soybean EST Project
 JOURNAL Contact: Shoemaker R/Public Soybean EST Project
 COMMENT Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3223 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Insert Length: 486 Std Error: 0.00.
 Location/Qualifiers

FEATURES

source

1..244
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl028-687"
 /clone_1lb="Gm-cl028"
 /tissue_type="roots of 'Superpod' plants"
 /lab_host="DH10B"
 /note="Vector: pBluescript II XR; site_1: EcoRI; site_2:
 XhoI; The mRNA was isolated from roots of Glycine max
 'Superpod' plants generously donated by Dr. Gary Stacey.
 The seedlings were inoculated with Bradyrhizobium
 japonicus, strain USDA110 prior to harvest. StrataGene's
 cDNA synthesis kit (catalog number 200401) was used to
 synthesize the cDNA. First-strand synthesis was performed
 with 5-methyl dCTP, hence the ligated cDNA was
 hemimethylated. A modification of StrataGene's
 first-strand synthesis primer was used. An 'anchor'
 nucleotide (V-A-C, or G) was added to the 3' end of the
 primer [GAGAGAGAGAGAGAGAGACTACTGTCGAG(T)18V] to anchor
 the primer at the 5' end of the poly(A) tract. After
 second-strand synthesis, the cDNA ends were filled in with
 cloned Pfu DNA polymerase, ligated to EcoRI adapters and
 subsequently phosphorylated. The XhoI site within the
 first-strand synthesis primer was then restricted by

digestion with XhoI; all XhoI sites in the cDNA would be
 protected by their hemimethylated status. The cDNA
 constructs were size-fractionated with a 500bp cutoff,
 using GibcoBRL Life Technologies' cDNA Size Fractionation
 column. The column eluent was then ligated into
 StrataGene's pBluescript II XR predigested vector
 (pBluescript II SK(+)) that has been digested with EcoRI
 and XhoI, and phosphorylated by StrataGene). Both the
 white and blue colonies appear to contain recombinant
 plasmids with cDNA inserts, based on size (n=25). This
 library was constructed by Dr. Paul Keim and Dr. Virginia
 Corryell."

BASE COUNT 65 a 37 c 65 g 77 t
 ORIGIN

Query Match 83.5%; Score 14.2; DB 20; Length 244;
 Best Local Similarity 76.5%; Pred. No. 7.6e+02;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATTTGGNTGGGNA 17
 ||:|:|:|:|:|:|:|:|:|
 Db 10 GACTTGGTTGGGGAA 26

RESULT 3

A1488429 247 bp mRNA EST 29-JUN-1999
 LOCUS A1488429/c
 DEFINITION EST246768 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
 cLED21013, mRNA sequence.

ACCESSION A1488429
 VERSION A1488429.1 GI:4383800
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum

Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta;
 Magnoliophyta: eudicotyledons: core eudicots: Asteridae: euasterids
 I; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 247)

REFERENCE Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
 Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
 C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
 S.D. and Giovannoni,J.
 Generation of ESTs from tomato carpel tissue
 Unpublished (1999)

TITLE Unpublished (1999)
 JOURNAL Contact: David Frisch
 COMMENT Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@CLEMSON.EDU.

Location/Qualifiers

FEATURES

source

1..247
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLED21013"
 /clone_1lb="tomato ovary, TAMU"
 /tissue_type="carpel"
 /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
 /lab_host="X11-Blue MRF"
 /note="Vector: pBluescript SK(-); site_1: EcoRI; site_2:
 XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
 directionally cloned cDNA in vector lambda Zap II with 5'
 and 3' ends located at the EcoRI and XhoI sites,
 respectively."

BASE COUNT 72 a 67 c 27 g 81 t
 ORIGIN

Query Match 83.5%; Score 14.2; DB 11; Length 247;
 Best Local Similarity 76.5%; Pred. No. 7.6e+02;

Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGGTGGGNA 17
 ||:||||| ||||| ||
 DB 222 GATTTGGTGGGNA 206

RESULT 4
 AZ015624 249 bp DNA GSS 25-FEB-2000
 LOCUS RPCT-23-258D7.TJ RPCT-23 Mus musculus genomic clone RPCT-23-258D7,
 DEFINITION DNA sequence.
 ACCESSION AZ015624
 VERSION AZ015624.1 GI:7091008
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 249)
 Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aknuret,
 B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
 and and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCT-23
 Unpublished (1999)
 Other_GSSs: RPCT-23-258D7.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@ligr.org
 Clones are derived from the mouse BAC library RPCT-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@edj.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.ligr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 258 row: D column: 7
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..249
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCT-23-258D7"
 /clone_1lb="RPCT-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pRACE3.6; Site: 1;
 Ecort: Site:2: Ecort: Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of Ecort and Ecort Methylase. Size
 selected DNA was cloned into the pRACE3.6 vector at the
 Ecort sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 87 a 54 c 39 g 69 t
 ORIGIN

Query Match 83.5%; Score 14.2; DB 113; Length 249;
 Best Local Similarity 76.5%; Pred. No. 7.6e+02;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGGTGGGNA 17
 ||:||||| ||||| ||
 DB 60 GATTTGGTGGGNA 76

RESULT 5
 AA650754

LOCUS AA650754 258 bp mRNA EST 31-OCT-1997
 DEFINITION 30868 Lambda-PRL2 Arabidopsis thaliana cDNA clone 281H6T7, mRNA
 sequence.
 ACCESSION AA650754
 VERSION AA650754.1 GI:2580846
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 258)
 Newman,T., debruin,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
 L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
 E. and Somerville,C.
 Genes galore: a summary of methods for accessing results from
 large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 Plant Physiol. 106, 1241-1255 (1994)
 95148729
 Contact: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
 Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 22313tcn@lm.cl.msu.edu
 Seq primer: T7 dye primer.

FEATURES
 source Location/Qualifiers
 1..258
 /organism="Arabidopsis thaliana"
 /strain="var columbica"
 /db_xref="taxon:3702"
 /clone="281H6T7"
 /clone_1lb="Lambda-PRL2"
 /note="Vector: lambda Zip-lox; Site:1: Sal; Site:2: Not;
 Lambda PRL2 is a cDNA library derived from equal
 quantiles of 4 pools of mRNA. The mRNA sources were 1) 7
 day germinated etiolated seedlings; 2) tissue culture
 grown roots; 3) staged plants half with 24 hour light
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
 same plants as 3 but aerial tissue (stems, flowers and
 siliques. The vector is BRL's lambda Zip-lox. The cDNA
 inserts were directionally cloned with Sal-Not arms using
 oligo dt primed cDNA."

BASE COUNT 74 a 42 c 62 g 70 t 10 others
 ORIGIN

Query Match 83.5%; Score 14.2; DB 5; Length 258;
 Best Local Similarity 76.5%; Pred. No. 7.6e+02;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGGTGGGNA 17
 ||:||||| ||||| ||
 DB 53 GATTTGGTGGGNA 69

RESULT 6
 AA651035 263 bp mRNA EST 05-JAN-1998
 LOCUS 31149 Lambda-PRL2 Arabidopsis thaliana cDNA clone 168A5XP 3', mRNA
 DEFINITION sequence.
 ACCESSION AA651035
 VERSION AA651035.1 GI:2581127
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 263)
 Newman,T., debruin,F.J., Green,P., Keegstra,K., Kende,H., McIntosh

TITLE
JOURNAL
MEDLINE
COMMENT

L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E., and Somerville, C.
Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones Plant Physiol. 106, 1241-1253 (1994)
95148729
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313c@lhm.cl.msu.edu
The sequence entry for this EST has been reverse complemented and is being submitted in the sense orientation.
Seq primer: M13-Universal.
Location/Qualifiers

FEATURES

source

1. 263
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="168A5X"
/note="Vector: lambda Z1P-lox; Site_1: Sal; Site_2: Not; lambda PR2 is a cDNA library derived from equal quantiles of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques. The vector is BRU's lambda Z1P-lox. The cDNA inserts were directionally cloned with Sal-Not arms using 0.1µg dt primed cDNA."

BASE COUNT

56 a 68 c 50 g 79 t 10 others
ORIGIN

Query Match 83.5%; Score 14.2; DB 5; Length 263;
Best Local Similarity 76.5%; Pred. No. 7.6e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTGGTGGGNA 17
11:11:1111111111
DB 90 GATTTGGATGGGCA 74

RESULT 7
LOCUS AI608251/c 278 bp mRNA EST 21-APR-1999
DEFINITION v193f09.x1 Knowles Solter mouse embryonic stem cell Mus musculus cDNA clone IMAGE:894569 3', mRNA sequence.
ACCESSION AI608251
VERSION AI608251.1 GI:4617418
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 278)
Marrin, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Materson, R., and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marrin M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:522529
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
Seq primer: primer name ambiguous
High quality sequence stop: 274.
Location/Qualifiers

FEATURES

source

1. 278
/organism="Mus musculus"
/strain="B6D F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:894569"
/note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(drf): 5'-CGGTGACCGTCGACGCTTTTCTTTT-3'. cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies)."
/dev_stage="embryo"
/lab_host="DH10B"
/db_host="DH10B"

BASE COUNT

57 a 75 c 48 g 98 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 11; Length 278;
Best Local Similarity 76.5%; Pred. No. 7.7e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTGGTGGGNA 17
11:11:1111111111
DB 99 GATTTGGCTGGGCA 83

RESULT 8
LOCUS BB264508/c 294 bp mRNA EST 07-JUL-2000
DEFINITION BB264508 RIKEN full-length enriched, 10 days neonate cortex Mus musculus cDNA clone AB30015G22 3', mRNA sequence.
ACCESSION BB264508
VERSION BB264508.1 GI:8960965
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 294)
Kono, H., Alizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Ode, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rctc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 324)
Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
A., Bolla, B., Marra, M., Hallier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers
Y., Person, B., Smaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shuh, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Other ESTs: sb39912.x1

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Glibco
High quality sequence stop: 311.

FEATURES

Location/Qualifiers
1..324

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl014-95"
/clone_1lb="Gm-cl014"
/tissue_type="Leaves, 2-3 week old seedlings, greenhouse
grown"
/lab_host="DH10B"

/note="Vector: pT73Pac (pT73, Pharmacia). Site_1: EcoRI;
Site_2: HindIII; This cDNA library was constructed from
mRNA isolated from leaves of 2-3 week old greenhouse grown
plants. Complementary DNA was synthesized from mRNA using
a 3' anchored poly (dT) primer. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by
directional ligation with EcoRI and HindIII. The cDNA fragments were
site of the pT73-Pac vector. The ligated cDNA fragments
were transformed into DH10B host cells (Gibco BRL). This
library was constructed by Dr. Randy Shoemaker and Dr.
John Expanding."

BASE COUNT 93 a 49 c 101 g 79 t 2 others

Query Match

Best Local Similarity 83.5%; Score 14.2; DB 13; Length 324;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATTTGCGTGGCGGNA 17

Db 192 GATTTGCGTGGCGGNA 208

RESULT 11

LOCUS A0034368/c

DEFINITION CIT-HSP-2319K4.TF CIT-HSP Homo sapiens genomic clone 2319K4, DNA
sequence.

ACCESSION A0034368

VERSION A0034368.1 GI:3300542

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 333)
Adams M.D., Rounsley S.D., Zhao S., Field C.E., Bass S., Linher K.,
Simon M., and Venter J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)

JOURNAL Unpublished (1998)

COMMENT Contact: Mark Adams

Department of Eukaryotic Genomics
The Institute for Genomic Research

FEATURES

source

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: maddams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends

Location/Qualifiers
1..333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2319K4"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 90 a 100 c 55 g 88 t

Query Match 83.5%; Score 14.2; DB 88; Length 333;
Best Local Similarity 76.5%; Pred. No. 7.8e+02;

Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATTTGCGTGGCGGNA 17

Db 237 GATTTGCGTGGCGGNA 221

RESULT 12

LOCUS BE460581/c

DEFINITION EST412000 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone cLE631N6, mRNA sequence.

ACCESSION BE460581

VERSION BE460581.1 GI:9504883

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eusterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 347)

AUTHORS Alcala J., Vredalov J., White R., van der Hoeven R.S., Holt I.E.,
Liang F., Hansen T.S., Craven M.B., Bowman C.L., Romling C.M.,
Nierman W., Fraser C.M., Martin G.B., Giovannoni J.J. and Tanksley
S.D.

Generation of ESTs from tomato fruit tissue, breaker stage
unpublished (2000)
Contact: David Fritsch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@CLEMSON.EDU

TITLE

JOURNAL

COMMENT

Unpublished (2000)
Contact: David Fritsch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@CLEMSON.EDU

FEATURES

source

Location/Qualifiers
1..347

/organism="Lycopersicon esculentum"
/cultivar="T496"

/db_xref="taxon:4081"

/clone="cLE631N6"

/clone_1lb="tomato breaker fruit, TIGR"

/tissue_type="Pericarp"

/dev_stage="breaker"

/lab_host="SOLR"

/note="Vector: pBluescriptSkncuadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopen accumulation on the blossom end of

the fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

BASE COUNT 120 a 71 c 40 g 116 t

Query Match 83.5%; Score 14.2; DB 35; Length 347;
Best Local Similarity 76.5%; Pred. No. 7.8e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATTTGGTGGGNA 17
Db 91 GATTTGGATGGGAA 75

RESULT 13
LOCUS AL378020 360 bp mRNA EST 03-AUG-2000
DEFINITION MCB35D12R1 MCBB Medicago truncatula cDNA clone MCB35D12 T7, mRNA sequence.
ACCESSION AL378020
VERSION AL378020.1 GI:9677772
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE 1 (bases 1 to 360)
Journet,E.P., Crespeau,H., van-Tulden,D., Gouzy,J., Jallion,O., Nibbel,A., Carreau,V., Chataigner,O., Kahn,D., Giamnazi,Pearson,V. and Gamas,P.

TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
JOURNAL Unpublished (2000)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES
source Location/Qualifiers
1..360
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone_1lb="MtrB35D12"
/clone_1lb="MtrB35D12"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
/note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using GigaPack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

BASE COUNT 128 a 33 c 80 g 119 t

Query Match 83.5%; Score 14.2; DB 14; Length 360;

Best Local Similarity 76.5%; Pred. No. 7.8e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATTTGGTGGGNA 17
Db 11 GATTTGGTGGGAA 27

RESULT 14
LOCUS AM350546/c 367 bp mRNA EST 01-FEB-2000
DEFINITION GM210009A10A2R Gm-r1021 Glycine max cDNA 3', mRNA sequence.
ACCESSION AM350546
VERSION AM350546.1 GI:6848256
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 367)
Vodkin,L., Kelm,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Expelding,J., Rahn,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics

University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888)919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES
source Location/Qualifiers
1..367
/organism="Glycine max"
/cultivar="Williams"
/db_xref="taxon:3847"
/clone_1lb="Gm-r1021"
/tissue_type="root"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Library Gm-r1021 is a sequence-driven, reracked set of the original library Gm-c1004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-c1004 library was constructed by Dr. Paul Kelm & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.kelmenau.edu, virginia.coryell@uau.edu. The cortig analysis to select unique genes was performed by the Laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota,
http://www.cbc.umn.edu/Research/Projects/Soybean/Index.html
Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois,
http://www.life.uiuc.edu/Dioechh/Keck.html."

BASE COUNT 115 a 94 c 59 g 10 others

Query Match 83.5%; Score 14.2; DB 21; Length 367;


```

QY      1 GAYTTGGNTGGGGNAA 17
          ||::||| ||| ||
Db      330 GACTTGGTTGGGGGAA 314

```

ACCESSION R63685 (HUMAN); mRNA sequence. *colga.53/43* FIBULIN-1, ISOFORM C PRECURSOR

SOURCE human.

REFERENCE AUTHORS

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Email: est@watson.wustl.edu
 Insert Size: 828
 High quality sequence stops: 192
 Source: IMAGE Consortium, LINT.
 This clone is available royalty-free through LINT; contact the
 IMAGE Consortium (info@image.lint.gov) for further information.
 Insert length: 828 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence stop: 192.

FEATURES
source

```

BASE COUNT
      89 a      92 c      127 g      63 t      18 others
constructed by Bento Soares and M.Fátima Bonaldo. "

/organism="Homo sapiens"
/db_xref="GDB:545291"
/db_xref="EAXON:9606"
/clone="IMAGE:138847"
/clone_lib="Soares placenta ND2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I ; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AATCGGAAGATTCGCGCCGACAGGAATTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bonaldo. "
```

Query Match	83.58;	Score 14.2;	DB 38;	Length 389;
Best Local Similarity	76.58;	Pred. No. 7.9e+02;		
Matches 13; Conservative	2;	Mismatches 2;	Indels 0;	Gaps 0;

OY	1	GATTTGCGTGGGNA	17
		:	
Db	351	GATTTTGGTTGGGAA	367

Search completed: October 20, 2000, 02:07:36
Job time: 1919 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 01:36:53 ; Search time 151.87 Seconds
(without alignments) 4212.508 Million cell updates/sec

Title: US-08-894-356C-1
1703

Perfect score: 1 TCATTATGAGCAAAATCCAA.....AAAAAAAAAAAAAAAAAAAA 1703

Sequence: 1 TCATTATGAGCAAAATCCAA.....AAAAAAAAAAAAAAAAAAAA 1703

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_36:*

1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:*

19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:*

20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:*

21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1698.2	99.7	1703	17	T37308 Aromatic acyl tran
2	202.4	11.9	1508	17	T37312 Aromatic acyl tran
3	193	11.3	1518	17	T37313 Aromatic acyl tran
4	186	10.9	1622	17	T37309 Aromatic acyl tran
5	172.2	10.1	1479	17	T37311 Aromatic acyl tran
6	39.8	2.3	3581	15	O44391 Aromatic acyl tran
7	39	2.3	340	20	X51737 Sequence of murine
8	39	2.3	1373	19	V52969 DNA encoding a hum
9	39	2.3	1701	21	V52969 Human G1 protein I
10	38.6	2.3	932	21	Z53336 Human secreted pro
11	38.4	2.3	2361	15	O68910 N. tabacum TOB-F12
12	38.4	2.3	2384	15	O68909 Human c-myc far up

13	38.2	2.2	1820	20	Z00434
14	38	2.2	1474	18	T90174
15	38	2.2	1557	20	V80671
16	38	2.2	4595	19	V59106
17	37.6	2.2	769	18	T79095
18	37.4	2.2	759	21	A02284
19	37.4	2.2	2400	18	T49364
20	37.2	2.2	988	13	O20500
21	37.2	2.2	2214	21	A26448
22	37.2	2.2	2227	21	A26365
23	37.2	2.2	605	21	Z80628
24	37	2.2	851	20	Z42177
25	36.8	2.2	1671	20	T97717
26	36.8	2.2	1383	21	T92614
27	36.8	2.2	3196	21	X36812
28	36.8	2.2	3721	20	X13348
29	36.6	2.1	596	21	Z88747
30	36.6	2.1	779	20	V84601
31	36.6	2.1	810	19	V63002
32	36.6	2.1	936	19	V63003
33	36.6	2.1	936	19	V62998
34	36.6	2.1	937	19	V62999
35	36.6	2.1	937	19	V63307
36	36.6	2.1	937	19	V62996
37	36.6	2.1	937	19	V62997
38	36.6	2.1	1029	19	V63308
39	36.6	2.1	1029	19	V63000
40	36.6	2.1	1633	20	X29700
41	36.6	2.1	1633	20	X29700
42	36.6	2.1	2144	19	V05530
43	36.6	2.1	2144	19	V05531
44	36.6	2.1	2780	21	A26422
45	36.6	2.1	3515	20	X60106

ALIGNMENTS

RESULT 1

T37308 T37308 standard; cDNA to mRNA: 1703 BP.

AC T37308;

DT 06-FEB-1997 (first entry)

DE Aromatic acyl transferase coding sequence.

KW Aromatic acyl transferase; transformation; anthocyanin pigment;

KW plants; acylation; colour; tone; colouration; colour change;

KW Gentiana triflora; Petunia hybrida; Petilla octinoides;

KW Senecio cruentus; Lavandula angustifolia; ds.

OS Gentiana triflora var. Japonica (Clone pGAT4).

FX Location/Qualifiers

FT Key 6.1415

FT CDS /tag= a /product= Aromatic acyl transferase.

XX MO9625500-A1.

XX 22-AUG-1996.

XX 16-FEB-1996; 96MO-JF00348.

XX 30-JAN-1996; 96JP-0046534.

XX 17-FEB-1995; 95JP-0067159.

XX 29-JUN-1995; 95JP-0196915.

XX (SUNR) SUNTORY LTD.

PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;

Human secreted pro
Oil seed rape cyst
Partial human DNAX
zebrafish dlifen
Full length human
Human colon cancer
Nicotiana glauca
Encodes fibrinogen
Human secreted pro
Human secreted pro
Human colon cancer
Human normal blood
Maize C-24 sterol
DNA encoding a bra
Enterococcus faeca
A. dichotoma Kabut
Human secreted pro
D. Immitis ankyrin
D. Immitis ankyrin
D. Immitis ankyrin
D. Immitis ankyrin
Nucleotide ndlank9
Nucleotide ndlank9
D. Immitis ankyrin
Nucleotide ndlank1
D. Immitis ankyrin
D. Immitis ankyrin
Human 7-transmembr
Rhesus monkey neur
Rhesus monkey neur
Human secreted pro
Arabidopsis thalia

PI Nakao M, Tanaka Y, Yonekura K;
XX WPI: 1996-393401/39.
DR P-PSDB: W047722.
PI DNA coding for aromatic acyl transferase - for transforming plants
PT which produce anthocyanin pigments and thus altering colour tone,
e.g. of flowers
XX
PS Claim 4: Page 53-57; 94pp; Japanese.
XX
CC Vectors containing DNA fragments encoding proteins of plant origin
CC with aromatic acyl transferase activity may be used to transform
CC plants which produce anthocyanin pigments. The aromatic acyl
CC transferase acylates the pigments in the flower resulting in colour
CC tone changes and allowing new colourations to be produced. Six
CC specific DNA sequences encoding aromatic acyl transferase from
CC different plants are described in T37308-T37313.
XX
SQ Sequence 1703 BP; 512 A; 353 C; 356 G; 482 T; 0 other;

Query Match 99.7%; Score 1698.2; DB 17; Length 1703;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1700; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TCATTATGAGCAAAATCCAAATGTAAGGTTCTGAAAATAGCCAAAGTTACACGCAT 60
DB 1 tcaataatgagcaaatccaaatgtaaggtctcgaaaaatgccaagtacacacacat 60
OY 61 CTGACACAACAGATGTGAGTATGCTACCGTAACGATTTCTGATATCCCTGGTTGC 120
DB 61 ctgacaacaacagatgtcgagttacgcgtacacgtatacatctcccttggtgc 120
OY 121 ACTGTAATAGATGAGTCCCTTCTGTTTACGACTTTCGTAACCAAGACATTTCT 180
DB 121 actgtaataagatgagtcctctctcttcttaccgacttccgtaaccaagaacacattct 180
OY 181 TGGACACTGTATCCCTAATCTTAAGGCCCTTTGCTCTCACTGAAAACACTACGTTTC 240
DB 181 tggacaactgtatccctaaatcctaagccctcttgcctcccttaaacactaccttc 240
OY 241 CGCTTAGCGGAAATTTGTTGATCCGATCAAAATCGGGCGAAAATGCCGAAGTTTCAGTACT 300
DB 241 cgcttagcggaatttgtgatcgcatcaaatcggcggaatgcgaagtttcagttact 300
OY 301 CCCGTAAGAGGCGACATCGATACTTGTATGCTTGCAGAGTGTACACGATTTTGACT 360
DB 301 cccgtaagagggcgactcgataacttgtatcgttcgagagtcgacagaatttgact 360
OY 361 ACCTTAAAGGTATCACTAGTATGATTCGATGATTTGATGGCGCTTTTATGTTATGC 420
DB 361 accttaaggtatcatactagatcgaatcccaatgattcgagccctttctatgtatgc 420
OY 421 CACGGGTTATAAGACCATGCAAGACTATAAGTATCCCGCTGAGCCGTGCAAGTAA 480
DB 421 cacgggtataaagacatgcaagactataaagtatcccgctgtagcgtgcaagtaa 480
OY 481 CCGTTTTCCTAACCGTGCCATAGCCGTGCTCTGACGGCACATCAATCAATTCAGATG 540
DB 481 ccgttttccctaaccgttgacagcgttgctctgacgscacatcatcatctgcagag 540
OY 541 CTAAGAAGTTTGTATGTCATCAATGCTTGGGCTATTTTAACAATTTGGGAAGAGC 600
DB 541 ctaagaagtttgtatgtcataatgcttgggctattttaacaatttgggaagagc 600
OY 601 CGGACTTGTTGTCGCGAATCTTCTTCATCTTTCATGATGAGATCAAGAATCTGT 660
DB 601 cggacttgtgtcgcgaaatctcttcacatcttcgatagatcgataatacaagatcgt 660
OY 661 ATGGCTTAGAGGAAACATTTTGAACGAAGTCAAGATGTTCTTGAAATGTTCTCTAGAT 720
DB 661 atggcttagaggaacattttagaacgaatgcaagatgtctctgaaatgtctctaagt 720

OY 721 TTGGAAGCAAAACCCCTCGATTCAACAAAGTACGAGCTACATATGTCCTTCCTTGCG 780
DB 721 ttggaagcaaaacccctcgattcaacaagtagcagctacataatgtctctcccttgccg 780
OY 781 AATCCAGAAGCTAAAGAAACAAAGTACTGAAATCTCAGAGATCCGACGCAATACGTG 840
DB 781 aatccagaagctaaagaaacaaagtactgaatctcagagatccgaaccgacatacgtg 840
OY 841 TTAACCACTTTCATATGACGTGTGATAGTATGACATGCTATGCTCAATTAAGAATG 900
DB 841 taacacactttcatatgacgtgtgatatgatacgtatgacatgcatggtcaataaagatg 900
OY 901 AGCTGTATCAAGAGATCATCATGCAAGCAGCAAAATGAGCTCCAGTACTTCAGTTTACG 960
DB 901 agctgtatcaagagatcatcatgcaagcagcaaaaatgagctccagtaacttcagttacg 960
OY 961 CGGATTTGCCGAGGACTTCTGACGCCCCCGTGTCCGCTTAATCACTTTGGCAACTGTCTTG 1020
DB 961 cggatttgccgagacttctgacgcccccgtgtccgcttaactcttggcaactgtcttg 1020
OY 1021 CGTCATGCGTTTCGAAAAGCAACACATTAAGAGTTAGTTGGGATTAAGGCTTCTTGTG 1080
DB 1021 cgtcatgcggttcgaaaagcaacacatTAAGAGttagttgggataaagggctcttg 1080
OY 1081 CAGTTGCACGTATTTGAGAGAGCCATTGAAAAGAGTTGCACAAAGGCGTTCTTG 1140
DB 1081 cagttgcacgtatTTGAGAGAGccattgaaaagagttgcacaaaggcgcttctg 1140
OY 1141 CAGATGCAAAACCTTGTTATCGGAATCTTAATGGAATCCCTTCAAAAAGATTTCTCGGGA 1200
DB 1141 cagatgcaaaaaccttgttatcggaatcttaatggaaatcccttcaaaaagatttccggga 1200
OY 1201 TTACGGGATCGGCTTAAGTTCGATGATGATGTTGATTTTGGATGGGAAAGCCTGCA 1260
DB 1201 ttacgggatcggcttaagtTCGATGATGATGTTGATTTTGGATGGGAAAGCCTGCA 1260
OY 1261 AACCTTGACATTCCTCTGTTGATTAAGCAATTCGATTAATGATTCGTCAGGAGATT 1320
DB 1261 aacttgacattccctctgttgatTAAGCAATTCGATTAATGATTCGTCAGGAGATT 1320
OY 1321 TTGAAAAGAGTGTGAGATTTGAGATATCTTAAGATTCATATGATGATTTTGCA 1380
DB 1321 ttgaaaagagtgtgagattTGAATatcatcttgcttaagttcataatgtagatctgcaa 1380
OY 1381 AATCTTTAAAGAAGCTTTTCTCTTGTGATAGTCTCTTTAATGAACATATTGCT 1440
DB 1381 aatctttgaagaagcttttctcttgatagctctcttcaatagaaccaatctgct 1440
OY 1441 GCAATTAATACCAAGTCCCTTTAGTAACTACACCAAAACCTTACTTTCGAGCGGGAAC 1500
DB 1441 gcaatTAATACCAAGTCCCTTTAGTAACTACACCAAAACCTTACTTTCGAGCGGGAAC 1500
OY 1501 ACCACAACGAGGTTCAATCACTAGAGAGTTGTACTTCATAAATTCAGAGGTGCAATATA 1560
DB 1501 accacaacgaggttcaatcatactagaaggtgtacttcaataatccagggcgcaatata 1560
OY 1561 CACCGTTGCTCTGTGAAAAGTTGAACCTTCACACCTGACATGCTGTTAGATAGTATGT 1620
DB 1561 cacggttgctctgtgaaaagttGAACCTTCACACCTGACATGCTGTTAGATAGTATGT 1620
OY 1621 AATAAGCCATTATATCTCCATTAAGATATCTATGCAATTAAGAAATGTTATGTTTA 1680
DB 1621 ataagccattatattctccataaagatatacctaagcaatagaacaatgtagtgc 1680
OY 1681 AAAAAAAAAAAAAAAAAAAAAA 1703
DB 1681 aaaaaaaaaaaaaaaaaaaaaa 1703

RESULT 2
T37312
ID T37312 standard; cDNA to mRNA; 1508 BP.

XX	T37312;
AC	
XX	
DT	06-FEB-1997 (first entry)
XX	
DE	Aromatic acyl transferase coding sequence.
XX	
KM	Aromatic acyl transferase; transformation; anthocyanin pigment;
KW	plants; acylation; colour; tone; coloration; colour change;
KM	Gentiana triflora; Petunia hybrida; Petilla octimoides;
KW	Scenecio cruentus; Lavandula angustifolia; ds.
XX	
OS	Senecio cruentus (Clone pcAr8).
XX	
FH	Key Location/Qualifiers
FT	CDS 3..1367
FT	/tag= a
FT	/product= Aromatic acyl transferase.
XX	
PN	MO9625500-A1.
XX	
PD	22-AUG-1996.
XX	
PF	16-FEB-1996; 96WO-JP00348.
XX	
PR	30-JAN-1996; 96JP-0046534.
XX	
PR	17-FEB-1995; 95JP-0067159.
XX	
PR	29-JUN-1995; 95JP-0196915.
XX	
PA	(SUNR) SUNTORY LTD.
XX	
PI	Ashikari T., Fujiwara H., Fukui Y., Kusumi N., Mizutani M.
XX	
PI	Nakao M., Tanaka Y., Yonekura K.
XX	
DR	WPI: 1996-393401/39.
XX	
P	P-PDB; W04726.
XX	
PT	DNA coding for aromatic acyl transferase - for transforming plants
PT	which produce anthocyanin pigments and thus altering colour tone,
PT	e.g. of flowers
XX	
PS	Claim 4; Page 69-72; 94pp; Japanese.
XX	
CC	Vectors containing DNA fragments encoding proteins of plant origin
CC	with aromatic acyl transferase activity may be used to transform
CC	plants which produce anthocyanin pigments. The aromatic acyl
CC	transferase acylates the pigments in the flower resulting in colour
CC	tone changes and allowing new colorations to be produced. Six
CC	specific DNA sequences encoding aromatic acyl transferase from
CC	different plants are described in T37308-T37313.
XX	
SQ	Sequence 1508 BP; 442 A; 294 C; 295 G; 477 T; 0 other;

Query Match	11.9%; Score 202.4; DB 17; Length 1508;
Best Local Similarity	52.0%; Pred. No. 11e-44;
Matches 720; Conservative 0; Mismatches 616; Indels 48; Gaps 10;	

OY	25	TGAAGCTTCGAAATGCCAAGTTACACCACTGCACACAACAGATGTGGATTAT	84
Db	1	tgaacattctcgaacatgcccgataatcgcccttcggagcca--tcgsgcatgcgt	57
OY	85	CGCTACCGGTAACTTTCGATATCCCTGGTTCGACTTGATAATAGATGAGTCCCTTC	144
Db	58	cgttactcttaactcttcggacattacttgctactcttcccttcggcaccatcttt	117
OY	145	TGTTTAGCACTTTCGATCCCAAGAACAATTTGAGACACTGTATTAACCTAATCTTA	204
Db	118	tcttctaagacttcacacattctaataccattcatcgaaactatttgttcccaggctaa	177
OY	205	AGGCGCTTTTGCTCCTCACTTAATAACACTAGCTTCCGCTTAAGCGAATTTGTTGATGC	264
Db	178	aacaactcttaacggtcaactcttaacaattttccgttttgctagttaattgatgtcat	237

Oy	265	CGATTCACA-----ATGGGGCGAAATGGCCAAATTCTCAGTACTCCCGTGATGAGCGC	31.15
Db	238	ttctctaaacctgaibggtctggggcttttaataaaaaaccagaataaaacaglttgaagtg	29.77
Oy	316	ACTCGAATCACTTGGATCGTTGGCGAGTCTGACCAAGATTTTGGACTTAAAGCTATC	37.55
Db	298	attctctgttggttaccttcttgagaagtgltgctgaccttaataatttgacggaaatc	35.7
Oy	376	AACCTGTACATTCACATGATTTTGCATGGCCCTTTTATGTTATGCCACGGGTTATAGGA	43.55
Db	358	atctctgaaaaagtgaaaaactttatccaactgtacacttcaatg---ggaatgcaatca	41.4
Oy	436	CCATGCAAGACATTAATGAATGATCCGCGCTGATAGCCGGTCAACGCTTTTTCATAC	49.5
Db	415	aattatgattctgctgaacggtcccaacttttcaactcnaagyaagtttttccggc	47.4
Oy	496	GTGGCATAGCCCTGGCTCTGACGGCAATCATTTCAATTGCAGATGCTMAAAGTTTGTAA	55.5
Db	475	cgggtatatacctaagaaatgaacatcatcaatgacctgtgtgaacgtagaacgcygtca	53.4
Oy	556	TGTTATCATCAATCTTGGGCGCTATATTAACAATTTTGGGAAGACCGCACTTGTGTCCG	61.5
Db	535	actctttgaagaaggtcggaactcgattatcaatccgtgtagatcggtctttttaaaga	59.4
Oy	616	CGAATCTTCTTCATCTTTTCGATAGATGATCAATCAAGAATCTGTATGGCCGTAGAGAA	67.5
Db	595	aaggaatctccacccggttttctgatatgtatgalttaaca-----tccacatttga	64.6
Oy	676	CATTTTGGACAGAAATGCAATGATTTCTTGAAATGTTCTGTAGATTTTGGAAAGCAACCC	73.5
Db	647	aaataaaglttgaaacatacaag---gctcgaaagtlttataaaacttcgagccctgtg	70.2
Oy	736	CTGTATTCACAAAGGTACGACCTACATATGTCCCTCTCTGTGTAATCCGAAGCTAA	79.5
Db	703	gtcccaactgaataaagtgctgcgaacglttgltgtgacccgaactaataatcaatcaactaa	76.2
Oy	796	AGAAACAAGTACTGAATCTCAGAGAGATCCACACCGACAAATACGTGTACGAGCTTCACA	85.5
Db	763	agaaaaaggtcttaa---cccgaatgcccacacttgagatacgtg---cacctttaa	81.6
Oy	856	TGACGTGTGATAGTACGTATGACATATGATGTGTCATAATCAAAAGATGACGTGTATCAGAG	91.5
Db	817	taactgtggttatatatatgtagtgcctagcgaaca-----ctcgtaaaaaatag	86.7
Oy	916	AATCATCGACGACGAAATAGAGCTGAGTACTTCAGTTTACAGCGGATTTGCCGAGNC	97.5
Db	868	gagaaagaagaagggagaaagagatctagaacagctataatcaacattgtgtcgatctc	92.7
Oy	976	TTTGACGCGCCCGCTGTCGCCCTTAACACTTGTGGCAACGTCTGTGGGTATACGTTGCCA	103.5
Db	928	gtcttgatcccaatcccaacagcccaactctgtgaactgtgtgacaaatgtgtccga	98.7
Oy	1036	AAGCAACACATTAAGAGTTAGTTGGGATTAAGAGGCTTCTTGTGCAGTTCAGAGTATTG	109.5
Db	988	ccataaaaaatgctgttttgactagcgaaaaatgtgtatgactgtgtgtctaagaacttg	104.7
Oy	1096	GAGAAAGCATTTGAAAGAGAGTTGCACACAGAAAGAGCGTTCTTGACAGTGCACAAACT	115.5
Db	1048	gagagcttatatgcaaaaatgatatataataaagaagcggaaatcttgaagaatgcgcgagat	110.7
Oy	1156	GGTTATCGGATCTTAATGGAATCCCTTCAAAAAAGATTTTCTGGGATTTACCGGATGGCCTA	121.5
Db	1108	gg---cagtaaccttcatatgatacccggtcaggaagatc---gggtctcgtgtaaaccta	116.1
Oy	1216	AGTTTCATTCGTATGCTGTAGATTTTGGATGGGGAAGCGCTGCACAAATTTGACATTACT	127.5
Db	1162	agccaactgtgacacttgtatcttgggtggggagagccgaataaaglatgagactgtct	122.1
Oy	1276	CTGTGTATTTGACAGATTAATGATTTTATGTATTCACGTCCAGGAGATTTTAAAGAGCTGCG	133.5
Db	1222	caatagactataatagctgataattctataataatgcaagcaaaaacatcagacaaagacttg	128.1

OY	1336	AGTTGAGTATCACTTCGCCCTACATTCCATTCGTGCATTTGCCAAATACTTTGAAGAAG	1393
Db	1282	aacttgattgagctacgcagatcgaatggaggcgctttcttagcatctgatgaga	1341
OY	1396	GCTT 1399	
Db	1342	gatt 1345	
RESULT	3		
T37313			
ID	T37313	standard; cDNA to mRNA; 1518 BP.	
XX			
AC	T37313;		
XX			
DT	06-FEB-1997	(first entry)	
XX			
DE	Aromatic acyl transferase coding sequence.		
XX			
KW	Aromatic acyl transferase; transformation; anthocyanin pigment;		
KV	plants: acylation: colour: tone: colouration: colour change;		
KW	Gentiana triflora; Petunia hybrida; Perilla ocimoides;		
KM	Senecio cruentus; Lavandula angustifolia; ds.		
XX			
OS	Lavandula angustifolia (clone PLAT21).		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1355	
FT		/+tag- a	
FT		/product= Aromatic acyl transferase.	
XX			
PN	W09625500-A1.		
XX			
PD	22-AUG-1996.		
XX			
PF	16-FEB-1996;	96MO-JP00348.	
XX			
PR	30-JAN-1996;	96JP-0046534.	
PR	17-FEB-1995;	95JP-0067159.	
PR	29-JUN-1995;	95JP-0196915.	
XX			
PA	(SUNR) SUNTORX LTD.		
XX			
PI	Asahikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;		
PI	Nakao M, Tanaka Y, Ionekura K;		
DR	WPI: 1996-393401/39.		
XX			
PRT	DNA coding for aromatic acyl transferase - for transforming plants		
PRT	which produce anthocyanin pigments and thus altering colour tone,		
PRT	e.g. of flowers		
XX			
PS	Claim 4; Page 73-76; 94pp; Japanese.		
XX			
CC	Vectors containing DNA fragments encoding proteins of plant origin		
CC	with aromatic acyl transferase activity may be used to transform		
CC	plants which produce anthocyanin pigments. The aromatic acyl		
CC	transferase acylates the pigments in the flower resulting in colour		
CC	tone changes and allowing new colourations to be produced. Six		
CC	specific DNA sequences encoding aromatic acyl transferase from		
CC	different plants are described in T37308-T37313. NOTE: This		
CC	sequence is supposed to cross reference with the protein described		
CC	in W04727, however there are so many discrepancies between the		
CC	polypeptide decoded from this sequence and the polypeptide given in		
CC	the specification and described in W04727 that the indexer decided		
CC	not to cross reference the two.		
XX			
Sequence	1518 BP; 384 A; 340 C; 366 G; 428 T; 0 other;		

Query Match	11.3%;	Score 193;	DB 17;	Length 1518;
Best Local Similarity	51.2%;	Pred. No. 3.8e-42;		
Matches 684;	Conservative 0;	Mismatches 605;	Indels 48;	Gaps 8

OY	69	ACAGATCTCGAGTTATGGCTAACCGGTAACATTTCTTGATATACCCCTGGTTGCACCTTGAAAT	128
Db	42	acggtggtgtagcagcgaactacccgctacacctctctgcacatgacgtggtgattctccac	101
OY	129	AAGATGACAGTCCCTTCTGTTTATAGCATCTTTCGTTACCAAGAACATATTTCTTGGACACT	188
Db	102	cccatgtctacgctctctctctcttaagaactccctglttccaaaccgctctctctcgaaac	161
OY	189	GTTATCCCTAATCTTAAAGCCCTCTTGTCTCTCACTCTAAACACTAGTCCGCTTAC	248
Db	162	gtcgtctccgaaactcaaacatccctctctctaacccccaacactctctcccccctta	221
OY	249	GGAAATTTGTTGATGCCGATCAAAATGGGGGAAATCCGAAGTTTACGATCTCCGCTAT	308
Db	222	tgcactctcaactcaaccctctatcgccgggaaaaagtcgagatctccgta-----tcag	275
OY	309	GAGGCGACATCGATACCTTTGATCGTTGGCCGAGCTCGACACGAGATTTTACTACCTTAA	368
Db	276	aacgttgatctcggtctcttcttaacgattatbgagtcgtcgtgaga---tcatcgaatcc	332
OY	369	GCTCATCACTGGTAGATTTCCAATGATTTTCATAGGCCCTTTTATGTTATGCCAGGGTT	428
Db	333	gtcataaatactactcgtcttgcctccatgagaatataatgaatctccaagtcgcgcgcg	392
OY	429	ATAAGGACATGACAGACTATAAGTATGATCCCGCTGGTACCGCTGCANATACCGTTT	488
Db	393	atagtcggagatctgtagtcggaatctgtctcaagtttagtcggtgaagtgatcctgltt	452
OY	489	CCTAACCGTGGCATAGCCGCTGGCTCTGACGGCACATCATTTCAATTGCAGATCTAAAGT	548
Db	453	cccggtcgcgggggtgtgcatctggaataacgaacacacacgcgtltagcgtatgcctatcg	512
OY	549	TTTGTAATGTTCAATCAATGCTGGGCCCTATATTAAACAATTTGGGAAAGCCGCGACTTG	608
Db	513	ttttagtggttataggaagatgtgggtctcatalcaactcaattggagagagaatgaatctc	572
OY	609	TTGTCCGCGA-----ATCTTCTTCATCTTTCGATAGATCGATTAATCAAGATCTGTAT	662
Db	573	tttgcgcgaaaaagtgaaatgtttgcggtttctgcacgactgcgtctgaaatataccgct	632
OY	663	GGCCATAGAGAAATATTTGGAGCAAAATCCAAAGATGTTCTTGAAATGTTCTCTAGATT	722
Db	633	aaattgacacatatctatgaaacaaacgcgcaaaaacgtc-----cgtltgaatcg	683
OY	723	GGAGCAAAACCCCTCATTTCAACAAGGTACGAGCTACATATGTCCTCTCCCTTCGTGA	782
Db	684	cagatcatalctttacacgaagatcgaatctcggtctaacctaactttcaaccaactciga	743
OY	783	ATCCAGAAGCTAAAGAACTAATCTGATATCTCAGAGATCCGAACCCGAATACGTGTA	842
Db	744	altaagaatatgaag-----ggtttgattcaagaagaagaccccaaatgtagtatacttc	797
OY	843	ACGAGCTTCACAAATGAGAGTGCTGTGATCGTATGATGACATGATGCTCAATCAAAATGAC	902
Db	798	tctctctcgtcgcgacatcgcaagtataatcttgacccggaatcgccaatc-----	847
OY	903	GTCGATCAGAGGAATCATGGAACGACGAAATAGACTGAGTACTTCACTTTACAGCG	962
Db	848	--ggtcgtagatctlaaagaagcgtgatacgacaacacgcgtcttcttctaactcgatc	905
OY	963	GATTGCGAGGACTTCTGAGCCCGCCGTCGCGCTACATACCTTTGGCAACGTCTGGG	1022
Db	906	gatttaagccgcggttcttgatccgcgcgcgcggggaaactctcgaaacgtctatcgcg	965
OY	1023	TCATGCGTTTGAAGACACACATATAAGATTTGTTGGGATTAAGAGCTCTTCTTTGCA	1082
Db	966	tttgcgtagtgaggaaagatccctgcgcgcggtatttgcgtgagatgaagggtgttctcgga	1025
OY	1083	GTTGACGCTATTGGAGAAACCATTTGAAAAGAGTTCACACAGAAAAGGCTTCTTGCA	1142
Db	1026	gtctaaagcgtatcgcggtgaatatagaataaagacg---agcgaacaaagatatctaaaga	1082

1999

[illegible]

Db	882	cttcaacgtagaactctgcaactcgtatctggaacatgcttgcgaaatcatatgaacacgcgcg	941
Qy	907	TATCAGAGGATCATCTGCACGACGAAATAGCTCGAGTACTTCAGTTTACAGCGGATT	966
Db	942	taagaagagaa---ggtgaaagagataaacaatgacgacaaacttaagtcttcaact	998
Qy	967	GCCGAGGACTTTCGACGCCCGCTGTCGGCTTAACACTTTGGCAACGTCTGGCTCAT	1026
Db	999	gcgcacacagtttctgcccgcgcgaatacccaaatcttggaaatcgacagtgccct	1058
Qy	1027	GCCTTGCAGAAAGCAACACATAAAGAGTTAGTTGGGGATMAAGGCTTCTTATGCAATTG	1086
Db	1059	gtatcgtcgtgatacgcaccatcagcaactcgttagaatagaagtgctgcgttagctgaa	1118
Qy	1087	CAGCTATTGGAGAACCCATTGAAAGAGGTGCACACGAAAGGCGTCTTGCAATG	1146
Db	1119	ccgcacatcgagatgctatccataagaagttcatcactgaagaatcttgaagagag	1178
Qy	1147	CAAAAACCTGGTTATCGGATCTA-----ATGGATCCCTTCAAAAAAGTTTCTCG	1197
Db	1179	atgtagatacgcgcgcgcgaatacactctgcgcgaacaaagtcgcagctcaattatctcg	1238
Qy	1198	GGATTACCGGATCGCCCTAAGTGTGATTGCGATAGGTGTAGATTTTTGGATGGGAAAGCCTG	1257
Db	1239	ttgatactcgcaacaagcaatgctgacatcttgaagcagatcttgctggggaagctg	1298
Qy	1258	CAAAATTTGACA---TTACCTCTGTTGATTATGCAGAAATGATTATGTTATGTAATGATCA	1314
Db	1299	aaaagcatgatactcgtttcaactaactccttcggcaacactaaatttgatactctcgtgccaa	1358
Qy	1315	GGGATTTTGGAAAAAGGTGTGGAGATTGAGATCATTTAGTCTTAAGATTGATGATGAT	1374
Db	1359	gaagatttaagaagcactctgagctcttgcatcttcttgctaagaatagatgltgacgat	1418
Qy	1375	TTGCAAAAATCTTT 1388	
Db	1419	ttgcacacatttt 1432	

XX	RESULT	5
XX	T37311	
ID	T37311 standard; cDNA to mRNA; 1479 BP.	
AC		
XX	T37311;	
DT		
DE	06-FEB-1997 (first entry)	
XX		
DE	Aromatic acyl transferase coding sequence.	
XX		
KW	Aromatic acyl transferase; transformation; anthocyanin pigment;	
KW	plants; acylation; colour; tone; colouration; colour change;	
KW	Gentiana triflora; Petunia hybrida; Perilla ocimoides;	
KW	Scenecio cruentus; Lavandula angustifolia; ds.	
XX		
OS	Perilla ocimoides (Clone pSAT208).	
XX		
FH	Key	Location/Qualifiers
FP	CDS	3..1343
FT		/tag= a
FT		/product= Aromatic acyl transferase.
XX		
PN	MO9625500-A1.	
PD		
PD	22-AUG-1996.	
XX		
PF	16-FEB-1996;	96WO-JP00348.
XX		
PR	30-JAN-1996;	96JP-0046534.
PR	17-FEB-1995;	95JP-0067159.
PR	29-JUN-1995;	95JP-0196515.
XX		
PA	(SUNR) SUNTORY LTD.	
XX		

PI Ashtikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
PI Nakao M, Tanaka Y, Yonekura K;
XX WPI: 1996-393401/39.
DR P-PSDB: W04725.
XX
XX DNA coding for aromatic acyl transferase - for transforming plants
PT which produce anthocyanin pigments and thus altering colour tone,
PT e.g. of flowers
XX
XX Claim 4; Page 65-69; 94pp; Japanese.
XX
XX Vectors containing DNA fragments encoding proteins of plant origin
CC with aromatic acyl transferase activity may be used to transform
CC plants which produce anthocyanin pigments. The aromatic acyl
CC transferase acylates the pigments in the flower resulting in colour
CC tone changes and allowing new colourations to be produced. Six
CC specific DNA sequences encoding aromatic acyl transferase from
CC different plants are described in T37308-T37313.
XX
XX Sequence 1479 BP; 420 A; 316 C; 331 G; 411 T; 1 other;

[illegible]


```

Db 690 ttaccgacgacccgcatcgcgaaccgcttcgtttccaccacccaatcaatgaatg 749
Oy 795 AAGAACAAAGTACTGATCTCAGAGATCCGAAACCGCAATAGCTGTAAGAGCTTACA 854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 750 aag-----ggtcgatccgctccagagctccaaagtttagtccatccctcatcttgta 803
Oy 855 ATGACGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 914
    || || || || || || || || || || || || || || || || || || || ||
Db 804 ggcattgcacgtctataatggtgctgcatacgaataatcattcaagaagtgag----- 858
Oy 915 GAATCATCGAACGACGAAATGAGCTCGAGTACTTCACTTTACACGCGATTGCCGAGA 974
    || || || || || || || || || || || || || || || || || || || ||
Db 859 -----accagaacagagatgcatcttcttctgctcgcgcgctcgaagcca 908
Oy 975 CTCTGACGCCCGCTGCTCGCCTAATCTTTGCAATGCTGCTGCTGCTGCTGCTGCTGCA 1034
    || || || || || || || || || || || || || || || || || || || ||
Db 909 cgattagatcccgctcctcgaataatcctcgggaactgctatctgctcgcgcgcg 968
Oy 1035 AAAGCAACACATAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1094
    || || || || || || || || || || || || || || || || || || || ||
Db 969 agaatgcgcgcgagcgagcgctggtgagagagaaggggtcttctgagctgagtgatc 1028
Oy 1095 GGAAGACCCATTGAAAAGAGTTGCACACGAAAGAGCGTTCTTGACGATGCAAAAAC 1154
    || || || || || || || || || || || || || || || || || || || ||
Db 1029 gcgcgagagataaaaaaagat---caacgacagaagataatagaacggtgagagaa 1085
Oy 1155 TGGTATCGGAATCTAATGCAATCCCTTCAAAAAGATTCTCGGGAATTACCGGATCGCCT 1214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1086 tggctcgcgagatctgtaaaagcgttcagaataatcattcttcggtgcagatcgaagc 1145
Oy 1215 AAGTTCGATCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1146 aagctaaactcttaagctgcaagatcttgatgagggagagagagagagagagagagatg 1205
Oy 1275 TCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1331
    || ||| || ||| || ||| || ||| || ||| || ||| || ||| || ||| ||
Db 1206 tcatgtgatgagagagagagagagagagagagagagagagagagagagagagagag 1265
Oy 1332 GTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1391
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 1266 ttggagagcttgctgctcctcgaagaagaagaatgcttctgctccttcttca 1325
Oy 1392 GAAGGCTTTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1451
    || || || || || || || || || || || || || || || || || || || ||
Db 1326 ctggagaaatagtgtaataatgtaataatgtaataatgtaataatgtaataatgta 1385
Oy 1452 CCAAGTCTTTAGTACACTACACCAACCTACTTT 1488
    |||| || || || || || || || || || || || || || || || || || ||
Db 1386 ttaagtgctgtaagctgtaagtaataatccctattat 1422

```

```

XX 25-AUG-1993: 93EP-0113602.
PF 28-AUG-1992: 92JP-0230028.
PR (FARH ) HOECHST JAPAN LTD.
PA Amann E, Kawal S, Okazaki M, Takeshita S, Tsujimura A;
XX WPI, 1994-076152/10.
DR P-PSDB: R49730.
XX New bone related, cadherin-like OSF-4 proteins - for treatment
PT and diagnosis of bone metabolic disease, and nucleic acid
XX encoding them
XX Claim 3: Page 13-17; 34pp: English.
XX CC CDNA libraries were constructed from the mouse osteoblastic cell
XX line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
XX then as much common DNA as possible removed by hybridisation between
XX the 2 libraries. Residual E1-specific DNA was amplified, inserted
XX into lambda gt10 and screened by plaque hybridisation. A minibank of
XX 273 E-specific clones was recovered, their inserts amplified and
XX used to screen total RNA from both cell types. One clone specific
XX for E1 was identified and sequenced. The insert from this clone was
XX used to screen cDNA prep. from E1 RNA and the longest posn. insert
XX cloned in pCEM 112f (+) to give pKOT164. This insert was sequenced;
XX it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
XX insert was also used to screen a cDNA bank prep. from human
XX CC osteosarcoma to identify 2 clones encoding the 2 human precursor
XX CC proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
XX CC respectively).
SQ Sequence 3581 BP; 1121 A; 774 C; 800 G; 886 T; 0 other:

Query Match 2.3%; Score 39.8; DB 15; Length 3581:
Best Local Similarity 79.7%; Pred. No. 0.58;
Matches 47; Conservative 0; Mismatches 12; Indels 0; Gaps 0:

Oy 1645 AAGTATCTATGCAATGAGACATGTTATGTTAAATGATGATGATGATGATGATGAT 1703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3469 aagcaaccttgaaataaaaaaagatcttcttaaaaaaataaaaaaataaaaaa 3527

RESULT 7
X51737
ID X51737 standard; DNA: 340 BP.
AC X51737;
DT 17-JUN-1999 (first entry)
DE DNA encoding a human secreted protein.
XX
XX Human secreted protein; cancer; immune disorder; infection;
XX inflammatory disorder; skin disorder; tumour; atherosclerosis;
XX restenosis; autoimmune disorder; Alzheimer's disease;
XX peripheral neuropathy; trauma; spinal cord injury; allergy;
XX hemiparetic disorder; skeletal disorder; neurological disorder;
XX arthritic disorder; asthma; immunodeficiency disease; AIDS;
XX transplant rejection; ss.
XX
XX Homo sapiens.
XX
XX WO9911293-A1.
XX 11-MAR-1999.
XX 03-SEP-1998; 98WO-US18360.
XX 12-SEP-1997; 97US-0058974.

```

PR	05-SEP-1997;	97US-0057626.	
PR	05-SEP-1997;	97US-0057663.	
PR	05-SEP-1997;	97US-0057669.	
PR	12-SEP-1997;	97US-0058666.	
PR	12-SEP-1997;	97US-0058667.	
PR	12-SEP-1997;	97US-0058973.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS;		
PI	Rosen GA, Ruben SM, Shi Y;		
XX			
DR	WPI: 1999-204988/17.		
XX	P-PSDB: Y12950.		
PS			
XX	Claim 1; Page 172; 215pp; English.		
CC	X51701-55 encode human secreted proteins. The polynucleotides and		
CC	their corresponding secreted polypeptides are useful for preventing,		
CC	treating or ameliorating medical conditions, e.g. by protein or gene		
CC	therapy. Pathological conditions can also be diagnosed by determining		
CC	the amount of the new polypeptides in a sample or by determining the		
CC	presence of mutations in the new polynucleotides. Specific uses are		
CC	described for each polynucleotide, based on which tissues they are		
CC	most highly expressed in, and include developing products for the		
CC	diagnosis or treatment of cancer, immune disorders, infection,		
CC	inflammatory disorders, skin disorders, tumours, atherosclerosis,		
CC	retenosis, autoimmune disorders, Alzheimer's disease, peripheral		
CC	neuropathies, trauma, spinal cord injuries, allergy, hematopoietic		
CC	disorders, skeletal disorders, neurological disorders, arthritic		
CC	disorders, asthma, immunodeficiency diseases, AIDS and transplant		
CC	rejection. The polypeptides are also useful for identifying their		
CC	binding partners.		
XX			
SO	Sequence 340 BP; 133 A; 69 C; 73 G; 65 T; 0 other;		
Query Match	2.3%; Score 39; DB 20; Length 340;		
Best Local Similarity	55.6%; Pred. NO. 0.35;		
Matches	75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;		
QY	1569 TCGTCTGAAAGATTGACCTCACCCTGACATGTTACGATAGCATTTGATAATGCC	1628	
Db	197 tctttgtactgaagaacctttaaggtggaaggtggaagagatgcttaataatca	256	
QY	1629 ATTATATCTTCGCATTAAGTATCCATATGCAATAGAGAACATCTTATNGTATAAAAAA	1688	
Db	257 attccaagcctccataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa	316	
QY	1689 AAAAAAAAAAAAAAAAAA 1703		
Db	317 aaaaaaaaaaaaaa 331		
RESULT	8		
V52969	V52969 standard; cDNA; 1373 BP.		
XX	V52969;		
AC			
XX			
DT	11-JAN-1999 (first entry)		
XX			
DE	Human G1 protein isoform beta (CASH-beta) cDNA.		
XX			
KM	G1 protein; CASH-beta; human; caapase homologue; Fas receptor;		
KW	modulator; apoptosis; cell death; inflammation; tumour; HIV;		
XX	therapy; ss.		
XX			
OS	Homo sapiens		

XX	Key	Location/Qualifiers
FH	482..1137	
FT	CDS	/tag- a
XX		
FN	MO9839435-A1.	
XX		
PD	11-SEP-1998.	
XX		
PF	26-FEB-1998;	98WO-TL00098.
PR	01-MAY-1997;	97IL-0120759.
PR	03-MAR-1997;	97IL-0120367.
XX		
PA	(YEDA) YEDA RES & DEV CO LTD.	
XX		
PL	Brodianski V, Goltssev Y, Kovalenko A, Varfolomeev E;	
PI	Wallach D;	
XX		
DR	WPI: 1998-495842/42.	
P	P-PSDB; W78904.	
PT	New DNA encoding isoforms of G1 protein which bind MORT-1 - and	
PT	regulate the effects of FAS and tumour necrosis factor receptors,	
PT	useful for killing of cells e.g. HIV and cancer cells	
XX		
XX	Claim 4; Fig 2; 132pp; English.	
CC	This CDNA sequence codes for the beta isoform (see W78904) of novel	
CC	human G1 protein. It was isolated from a human skin fibroblast	
CC	cDNA library using a cDNA probe corresponding to the G1 sequence.	
CC	G1-beta (also called CASH beta, CASH being caspase homologue)	
CC	and a longer isoform, G1-alpha (see W78903), represent 2 splice	
CC	variants of the G1 protein. These G1 proteins are capable of	
CC	binding to, or interacting directly or indirectly, with MORT-1	
CC	or with MORT-binding proteins such as Mch4 (CASP-10) and MACH	
CC	(CASP-8), and thereby of binding to the intracellular domain of	
CC	the FAS-R receptor, to which MORT-1 binds, or of binding to the	
CC	intracellular domain of the p35 tumour necrosis factor (TNF)	
CC	receptor, to which TRADD binds and to which TRADD protein MORT-	
CC	1 binds. Hence, they are considered as mediators or modulators of	
CC	FAS-R having a role in e.g. the signalling process that is initiated	
CC	by the binding of FAS ligand to FAS-R, and also having a role in	
CC	the signalling process initiated by the binding of TNF to p55-R.	
CC	G1 DNA (I) and polypeptide (II), vectors and fragments are used to	
CC	regulate cell death or inflammatory processes. (II) is used to	
CC	inhibit cell death, and its inhibitors augment/enhance the	
CC	processes. (I) and (II) regulate the FAS-R ligand or TNF effect on	
CC	cells carrying an FAS-R or p55-R. Tumour, HIV-infected or other	
CC	diseased cells can be treated using a viral vector encoding a viral	
CC	surface protein, which binds a specific cell surface receptor and a	
CC	sequence encoding (II), which kills the cell. Antisense	
CC	oligonucleotides, introduced using the above vector, block the	
CC	expression of (II) and can also regulate the above effects. These	
CC	effects can also be regulated using a vector encoding a ribozyme	
CC	that interacts with a cellular mRNA encoding (II), and allows (II)	
CC	expression.	
XX		
SO	Sequence 1373 BP; 417 A; 283 C; 335 G; 338 T; 0 other;	
Query Match	2.3%; Score 39; DB 19; Length 1373;	
Best Local Similarity	68.4%; Pred. No. 0.63;	
Matches 54; Conservative 0; Mismatches 25; Indels 0; Gaps 0;		
DY	1625 TGCCATTAATACTCCATAAAGTTCCTATGCAATAGAGAACATGTATATGTTAAAAA	1684
Db	1258 tccaatcttctgtactactataatgctaataaataataatcctgttactctttaaaa	1317
DY	1685 AAAAAAAAAAAAAAAA 1703	
Db	1318 aaaaaaaaaaaaaaa 1336	

RESULT	9	
ID	233336	standard; CDNA; 1701 BP.
XX	AC	233336;
XX		
DT	29-FEB-2000	(first entry)
XX		
DE		Human secreted protein clone pu261_1 nucleotide sequence SEQ ID NO:41.
KW		Human; secreted protein; nutritional; cytokine; cell proliferation;
KW		differentiation; immune stimulating; vaccine; suppression;
KW		hematopoiesis regulation; tissue growth; activin; inhibin;
KW		chemotactic; chemokine; haemostatic; thrombolytic; receptor;
KW		ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
XX		tumour inhibition; gene therapy; ss.
OS		Homo sapiens.
XX		
PN		WO957132-A1.
XX		
PD		11-NOV-1999.
XX		
PF		07-MAY-1999; 99WO-US09970.
XX		
PR		07-MAY-1998; 98US-0084564.
PR		02-JUN-1998; 98US-0087645.
PR		22-JUL-1998; 98US-0093712.
PR		31-JUL-1998; 98US-0094935.
PR		10-AUG-1998; 98US-0095880.
PR		11-AUG-1998; 98US-0096068.
PR		06-MAY-1999; 99US-0096068.
XX		
PA		(GENEY) GENETICS INST INC.
XX		
PI		Mebers K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI		Jefferys D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;
PI		Diblasio-Smith E, Wildom A;
XX		
DR		WPI: 2000-052937/04.
XX		P-PSDB; Y53018.
PT		New polynucleotides encoding secreted human proteins, derived from
XX		adult placenta, adult retina, fetal brain, fetal
XX		
PS		Claim 50: Page 400; 492pp; English.
XX		
CC		The present invention describes new human secreted proteins which were
CC		isolated from adult placenta, adult retina, foetal brain, foetal kidney,
CC		adult blood, adult brain, adult thyroid, adult bladder, adult neural
CC		tissue, adult testes, and adult lymph node cDNA libraries. The human
CC		secreted proteins, and the polynucleotides encoding them, are predicted
CC		to have biological activities which would make them suitable for
CC		treating, preventing or ameliorating medical conditions in humans and
CC		animals. Suggested activities include nutritional activity, cytokine
CC		and cell proliferation/differentiation activity, immune stimulating
CC		(e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC		activity, tissue growth activity, activin/inhibin activity,
CC		chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC		activity, receptor/ligand activity, anti-inflammatory activity,
CC		adherin/tumour invasion suppressor activity, and tumour inhibition
CC		activity. The polynucleotides are also stated to be useful for gene
CC		therapy. 233316 to 233373 encode human secreted proteins, and Y52998 to
XX		Y53060 represent human secreted proteins, given in the present invention.
XX		
XX		Sequence 1701 BP; 382 A; 456 C; 494 G; 369 T; 0 other;

Query Match 2.3%; Score 39; DB 21; Length 1701;
Best Local Similarity 55.6%; Pred. No. 0.69;
Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Oy	1569 TCCCTCGAAAGATTGACCTCACACTGCATGGTGTTACGATAGGTATTTGATAAATGCC	1628
Dd		1599
Oy	1639 ATTATATACCTTCATTAAGTATCTCATGCATAGACACATCTTTATGTTTTAAAAAAA	1688
Dd		1659
Oy	1600 attcacaagctcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	1659
Dd	1689 AAAAAAAAAAAAAAAA 1703 	1674
RESULT	10	
ID	Z51920 standard; DNA: 932 BP.	
XX	Z51920;	
XX	04-JUL-2000 (first entry)	
N.	tabacum TOB-F12 gene for controlling wild fire disease.	
KW	TOB-F12 gene: wild fire disease; tobacco; resistance gene; R gene; non-host disease resistance; plant pathogen; hypersensitive response; HR; Phytophthora infestans elicitor; INF1; transgenic plants; viral pathogen; fungal pathogen; bacterial pathogen; nematodal pathogen; Pytophthora infestans; ss.	
OS	Nicotiana tabacum.	
FH	Key	Location/Qualifiers
FT	CDS	99..473 /*tag= a /product= "TOB-F12 protein"
PX	MO200012736-A2.	
PD	09-MAR-2000.	
PX	31-AUG-1999:	99MO-US19899.
PX	31-AUG-1998:	98US-0098402. (MONS) MONSANTO CO.
PA	Rommens CMT, Swords KMM, Yan H, Zhang B; MPI: 2000-256651/22. P-PSDB; Y70550.	
PT	Identification of non-host plant disease resistance genes comprises expressing resistance and non-host inducible genes in susceptible plants -	
Claim 14;	Page 90; 94pp; English.	
The patent discloses a method for identifying genes that enhance levels of disease resistance if expressed in susceptible plants. The method is useful for isolating disease resistance genes (R gene) in plants. These genes confer non-host disease resistance to plants by responding to hypersensitive response (HR) in plant pathogens. The R-genes identified trigger a defense reaction in tobacco that is dependent on the presence of the Phytophthora infestans elicitor INF1. The genes are useful for generating pathogen-resistant transgenic plants. They can be used to control viral, fungal, bacterial or nematodal pathogens, e.g., Phytophthora, Erisiphe and Puccinia. The present sequence is TOB-F12 gene encoding a homologue of the 21kDa protein of Daucus carota. Expression of this gene was shown to partially control wild fire disease of tobacco.		

CC Insect pests and fungal diseases, improved yield and/or quality of
CC harvested product, and novel mechanisms for the control of plant
CC fertility.
XX
S0 Sequence 1474 BP; 528 A; 265 C; 316 G; 365 T; 0 other;

SQ Sequence 1474 BP; 528 A; 265 C; 316 G; 365 T; 0 other;

Query Match	2.28;	Score 38;	DB 18;	Length 1474;
Best Local Similarly	60.88;	Pred. No. 1.2;		
Matches 62; Conservative	0;	Mismatches 40;	Indels 0;	Gaps 0;

Oy 1602 GTGTTCACGATAGCTATTGGTATAATGCCATTATATACTTCACATAAAGTATCCTTATGCMAATA 1661

Db 1338 gtcttaagaagatataataataaatatataattcttctaataaaaaaaaaaaaaaa 1397

Oy	1662	GAGACATGTTATGTGTTAAAAA	AAAAAAAAAAAAAAAAAA	1703
Db	1398	aaaaaaaaaaaaaaaaaaaaaa	aaaaaaaaaaaaaaaaaaaa	1439

RESULT	15
V80671	
ID	V80671 standard; cDNA; 1557 BP

DT 15-MAR-1999 (first entry)

DE Partial human DNAX toll-like receptor DTLR7 encoding cDNA.

KM DNAX toll-like receptor; DTLR; Drosophila toll receptor; IL-1 receptor
 KW Interleukin 1 receptor; phosphate metabolism; innate immunity response
 KM modulate inflammatory function; morphological effect;
 KW immunological disorder; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	CDS	1..516

```

/ product= "partial DTLR7"
FT

```

FT	/*tag=	b
FT	/note=	"nucleotide 278 designated G, may be G or C"
FT	misc_feature	445

FT /note= "nucleotide 445 designated A, may be A or T"

	/tag= d
FT	/note= "nucleotide 572 designated C, may be A, C, G or T"
FT	
FT	
misc_feature	593

/note= "nucleotide 593 designated C, may be A, C, G or T"

```

FT      /*tag= f
FT      /note= "nucleotide 600 designated C, may be A, C, G or T
FT      misc_feature 607

```

FT /note- "nucleotide 607 designated C, may be A, C, G or T"

FT	/*tag=	h
FT	/note=	"nucleotide 617 designated C, may be A, C, G or T
FT	misc_feature	622

FT /note= "nucleotide 622 designated C, may be A, C, G or T"

FT	/*tag= j
FT	/note= "nucleotide 625 designated C, may be A, C, G or T
FT	misc_feature
	631

FT /note= "nucleotide 631 designated C, may be A, C, G or T"

13

FT	/note="nucleotide 640 designated C, may be A, C, G or T"
FT	
misc_feature	646

FT /note= "nucleotide 646 designated C, may be A, C, G or T"

FT	/*tag=	n
FT	/note=	"nucleotide 653 designated C, may be A, C, G or T"
FT	misc_feature	719

FT /note= "nucleotide 719 designated C, may be A, C, G or T"

FT	/*tag="	p
FT	/note="	nucleotide 775 designated C, may be A, C, G or T"
FT	misc_feature	861

FT /note= "nucleotide 861 designated C, may be A, C, G or T"

PN W09850547-A2

PD 12-NOV-1998.

PF 07-MAY-1998; 98WO-US08979.

PR 05-MAR-1998; 98US-0076947.

PR 22-JAN-1998; 98US-0072212.

PA (SCHE) SCHERING CORP.

PI Bazan JF, Hardiman GT, Kastelein RA, Rock FL

DR WPI; 1999-059670/05

XXII

PT phosphate metabolism, modulate inflammatory function or innate

[illegible]

The present invention specifically describes human DNAX toll-like receptors 2 to 10 (DILR2-10). The present sequence encodes partial human DILR2 given in the present invention. Also described are: (1) a fusion protein comprising a DILR protein or peptide; (2) a binding compound, preferably an antibody or antibody fragment which specifically binds to a DILR protein or peptide; (3) a nucleic acid encoding a DILR protein or peptide; (4) an expression vector comprising the nucleic acid of (3); and (5) a host cell comprising the vector of (4). The host cell of (5) can be used to produce the DILR proteins. The DILR proteins can be used to alter phosphate metabolism, to modulate inflammatory function, innate immunity responses or morphological effects. The DILR proteins can be used in the treatment of conditions exhibiting abnormal expression of the receptors of their ligands. These abnormalities are typically manifested by immunological disorders.

Sequence 1557 BP; 459 A; 316 C; 319 G; 463 T; 0 other;

Query Match	2.28;	Score 38;	DB 20;	Length 1557;
Best Local Similarity	65.18;	Pred. No. 1.2;		
Matches 56; Conservative	0;	Mismatches 30;	Indels 0;	Gaps 0;

QY 1618 TGTTCATGCGATTATATACCTCCATAAGTATCCTATGCAAATAGAGAACAAGTTATGTG 1677

Db 1471 Tctataattttaattgcacattgctctaactttaaatgaatgaatataaaatggttcaatt 1530

QY	1678	TTAAAAAAAAAAAAAAAAAAAAA	1703
Db	1531	ttaaaaaaaaaaaaaaaaaaaaa	1556

Sun Oct 22 12:45:49 2000

us-08-894-356c-1.rng

Page 13

Job time: 5857 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 01:36:02 ; Search time 3710.09 Seconds
(without alignments)

2004.800 Million cell updates/sec

Title: US-08-894-356C-1

Perfect score: 1703

Sequence: 1 TCATTATGAGCAATCCAA.....AAAAAAAAAAAAAAAAAAAA 1703

Scoring table: IDENTIFY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_com:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_com:*
20: em_ov:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_v1:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_da1:*
35: em_da2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pr4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pr5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_v12:*
73: gb_da3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: em_htg11:*
78: em_htg12:*
79: em_htg13:*
80: em_htg14:*
81: em_htg15:*
82: em_htg16:*
83: em_htg17:*
84: em_htg18:*
85: em_htg19:*
86: em_htg20:*
87: em_htg21:*
88: em_htg22:*
89: em_htg23:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_sts1:*
93: gb_sts2:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1703	100.0	1703	5	E12753
2	1679	98.6	1679	5	AB010708
3	204	12.0	1508	5	E12757
4	189.2	11.1	1622	5	E12754
5	189.2	11.1	1622	5	AB026494
6	172.2	10.1	1479	5	E12756
7	171.4	10.1	1476	7	AB029340
8	114	6.7	119814	7	AC003027
9	109.2	6.4	101176	7	AC002560
C 10	71.8	4.2	82360	8	AP000606
C 11	71.8	4.2	166884	60	AC074226
12	64.2	3.8	66237	7	AB016892

C	13	63	3.7	136047	59	AC069470	AC069470 Arabidops
C	14	61.8	3.6	85690	7	AB028618	AB028618 Arabidops
C	15	61.8	3.6	136047	59	AC069470	AC069470 Arabidops
C	16	51.2	3.0	85690	7	AB028618	AB028618 Arabidops
C	17	45.6	2.7	321003	71	EFMAIAP3	AL035476 Plasmid
C	18	44.4	2.6	269	57	EC4270224	AL035476 Plasmid
C	19	44	2.6	46739	8	AP002052	AB006696 Arabidops
C	20	44	2.6	78379	7	AB006696	AB006696 Arabidops
C	21	43.4	2.5	117729	59	AC069169	AC069169 Homo sapi
C	22	43	2.5	161891	30	AC008206	AC008206 Drosophila
C	23	42.8	2.5	3314	33	AF163834	AF163834 Dictyoste
C	24	42.8	2.5	188925	51	AC022050	AC022050 Homo sapi
C	25	42.4	2.5	144334	30	AC004153	AC004153 Plasmid
C	26	42.4	2.5	154425	31	AC008911	AC008911 Homo sapi
C	27	42	2.5	170920	38	CNS01DXJ	AL139353 Human chr
C	28	42	2.5	202496	38	CNS01DM6	AL136418 Human chr
C	29	42	2.5	202496	38	CNS01DX6	AL139054 Human chr
C	30	42	2.5	219936	42	AC016299	AC016299 Homo sapi
C	31	41.8	2.5	2222	92	HOMERLIN	AL11284 Homosapiens
C	32	41.8	2.5	200543	42	AC016726	AC016726 Homo sapi
C	33	41.8	2.5	207702	68	AL354820	AL354820 Homo sapi
C	34	41.6	2.4	356	48	PSNOD6MR	X63699 P. sativum P
C	35	41.2	2.4	14955	32	AF000580	AF000580 Dictyoste
C	36	41.2	2.4	100000	38	AP000096	AP000096 Homo sapi
C	37	41.2	2.4	100000	38	AP000200	AP000200 Homo sapi
C	38	41.2	2.4	160319	54	AC026505	AC026505 Homo sapi
C	39	41.2	2.4	176099	38	AP000240	AP000240 Homo sapi
C	40	41.2	2.4	340000	38	AP001705	AP001705 Homo sapi
C	41	41	2.4	1698	92	IKO668307	AL109118 Homo sapi
C	42	41	2.4	4572	11	AK000465	AK000465 Homo sapi
C	43	41	2.4	80318	42	AC016300	AC016300 Homo sapi
C	44	41	2.4	205495	51	AC022509	AC022509 Homo sapi
C	45	40.8	2.4	1019	8	AV070215	U70215 Amaranthus

ALIGNMENTS

RESULT	1	LOCUS	E12753	1703 bp	DNA	PAT	24-JUN-1998
DEFINITION			Gentianatripliflora mrna	acetyltransferase, complete	cds.		
ACCESSION			E12753				
VERSION			E12753.1	GI:3251585			
KEYWORDS			JP 1997070290-A/1.				
SOURCE			unidentified.				
ORGANISM			unclassified.				
REFERENCE			1 (bases 1 to 1703)				
AUTHORS			Ashikari, T., Tanaka, Y., Fujiwara, H., Nakao, M., Fukui, Y., Yonekura, K., Mizutani, M., and Kusumi, T.				
TITLE			GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY				
JOURNAL			Patent: JP 1997070290-A 1 18-MAR-1997;				
COMMENT			SUNORY LTD				
OS			Gentianatripliflora				
PN			JP 1997070290-A/1				
PD			18-MAR-1997				
PE			30-JAN-1996 JP 1996046534				
PR			17-FEB-1995 JP 95P 67159, 29-JUN-1995 JP 95P 196915				
PI			ASHIKARI TOSHIHIKO, TANAKA YOSHITAKU, FUJIMURA HIROYUKI, PI NAKAO MASAHITO, FUKUI YUKO, YONEKURA KEIKO, MIZUTANI MASAKO, KUSUMI TAKAKI				
PC			C12N15/09, A01H1/04, C07H21/04, C07K14/42, C12N9/10, (C12N9/10, PC C12R1:865),				
CC			Strandedness: Double;				
CC			Topology: Linear;				
CC			key				
CC			Location/Qualifiers				
CC			source				
CC			1. 1703				
CC			/organism='Gentianatripliflora'				
CC			/variety='japonica'				
CC			/tissue_type='petal'				

```

Oy 841 TAACGACGTTACACATGACGTGTGATACGTATGACATGTCATGTCATCAATCAAAAGATG 900
    |||||||
Db 841 TAACGACGTTACACATGACGTGTGATACGTATGACATGTCATGTCATCAATCAAAAGATG 900
Oy 901 ACCTGATATCAGAGGATATCATCGACAGCAAGAAAATGAGTCTGAGTACTTCAGTTTACAG 960
    |||||||
Db 901 ACCTGATATCAGAGGATATCATCGACAGCAAGAAAATGAGTCTGAGTACTTCAGTTTACAG 960
Oy 961 CGGATGCGAGAGCACTTTCAGAGCCCCCGGTCGCCCTACACTCTTTGGCAACTGCTTG 1020
    |||||||
Db 961 CGGATGCGAGAGCACTTTCAGAGCCCCCGGTCGCCCTACACTCTTTGGCAACTGCTTG 1020
Oy 1021 CGTATCGCGTTCGCAAAAGCAACATTAAGAGTTAGTGGGATTAAGGGCTTCTTGTTG 1080
    |||||||
Db 1021 CGTATCGCGTTCGCAAAAGCAACATTAAGAGTTAGTGGGATTAAGGGCTTCTTGTTG 1080
Oy 1081 CAGTTCAGCTATTTGAGAGAACCAATGAAAAGGTTGCAACAAGAAAAGGCGTTCTTG 1140
    |||||||
Db 1081 CAGTTCAGCTATTTGAGAGAACCAATGAAAAGGTTGCAACAAGAAAAGGCGTTCTTG 1140
Oy 1141 CAGATGCAAAAACCTGTTATCGGAATCTATGGAATCCCTTCAAAAAGATTTCTGGGA 1200
    |||||||
Db 1141 CAGATGCAAAAACCTGTTATCGGAATCTATGGAATCCCTTCAAAAAGATTTCTGGGA 1200
Oy 1201 TTACCGGATCGCTTAAGTTGATGATGCTAGATTTGGATGGGGAAGCGTCGAA 1260
    |||||||
Db 1201 TTACCGGATCGCTTAAGTTGATGATGCTAGATTTGGATGGGGAAGCGTCGAA 1260
Oy 1261 AATTTGACATTTACCTCTGTTGATTAAGCAATTTGATTTATGATTCAGTCCAGGGATT 1320
    |||||||
Db 1261 AATTTGACATTTACCTCTGTTGATTAAGCAATTTGATTTATGATTCAGTCCAGGGATT 1320
Oy 1321 TTGAAAAAGGTGTGGAGATTTGAGATTCATTCCTTAAGATTCATATGATGATTCGAA 1380
    |||||||
Db 1321 TTGAAAAAGGTGTGGAGATTTGAGATTCATTCCTTAAGATTCATATGATGATTCGAA 1380
Oy 1381 AAATCTTTGAAGAAGGCTTTGCTTTGTCATAGTCTCTTAATGAACATATTGCT 1440
    |||||||
Db 1381 AAATCTTTGAAGAAGGCTTTGCTTTGTCATAGTCTCTTAATGAACATATTGCT 1440
Oy 1441 GCAATTAAGTACCAAGTCTTTAGTAACTACACCAAAACCTTCTGAGGCGGGAAC 1500
    |||||||
Db 1441 GCAATTAAGTACCAAGTCTTTAGTAACTACACCAAAACCTTCTGAGGCGGGAAC 1500
Oy 1501 ACCACAACGAGGTTCACTCTGAGAGGTTGCTTCTATTAATCCAGAGGTGAATATA 1560
    |||||||
Db 1501 ACCACAACGAGGTTCACTCTGAGAGGTTGCTTCTATTAATCCAGAGGTGAATATA 1560
Oy 1561 CACCGTTGCTCTGAAAAGTTGACCTCAGCTGACATGCTTACGATAGTATTGT 1620
    |||||||
Db 1561 CACCGTTGCTCTGAAAAGTTGACCTCAGCTGACATGCTTACGATAGTATTGT 1620
Oy 1621 ATAAATGCCATTAATATCTTCATTAAGTATCTATGCAATAGAGAATGTTATGTGTTA 1680
    |||||||
Db 1621 ATAAATGCCATTAATATCTTCATTAAGTATCTATGCAATAGAGAATGTTATGTGTTA 1680
Oy 1681 AAAAAAAAAAAAAAAAAAAAAA 1703
    |||||||
Db 1681 AAAAAAAAAAAAAAAAAAAAAA 1703

```

```

REFERENCE
AUTHORS      1 (bases 1 to 1679)
TITLE        Fujiwara, H., Tanaka, Y. and Kusumi, T.
JOURNAL      Direct Submission
              Submitted (22-JAN-1998) to the DDBJ/EMBL/GenBank databases.
              Hiroyuki Fujiwara, Suntory LTD., Inst. of Fundamental Research;
              1-1-1, Wakayamada, Shiamoto-cho, Osaka 618-0024, Japan
              (E-mail: Hiroyuki.Fujiwara@suntory.co.jp, Tel: +81-75-962-8807,
              Fax: +81-75-962-8262)
REFERENCE    2 (sites)
AUTHORS      Fujiwara, H., Tanaka, Y., Yonekura-Sakakibara, K.,
              Fukuchi-Mizutani, M., Nakao, M., Fukui, Y., Yamaguchi, M., Ashikari, T.,
              and Kusumi, T.
TITLE        cDNA cloning, gene expression and subcellular localization of
              anthocyanin 5-aromatic acyltransferase from Gentiana triflora
              Plant J. 16 (4), 421-431 (1998)
JOURNAL      99097837
MEDLINE
FEATURES
source       Location/Qualifiers
              1..1679
              /organism="Gentiana triflora"
              /db_xref="taxon:55190"
              /clone="pgAT4"
              /tissue_type="petal"
              6..1415
              /codon_start=1
              /product="anthocyanin 5-aromatic acyltransferase"
              /protein_id="BA074428.1"
              /db_xref="GI:4185599"
              /translation="MEQIOMNVLEKSCQVTPPSDTTDELSPVTFPDIPLHLKMQ
              SLTFDEPFRPHFLDPTVPMNKASLSLTKHYVLSGLMLPRLSGEMPKROYRDE
              GSDTLIVAESDPRDYLKHOIUDNSDHLGFLPMPRIPIRMOXYKXPLVAVQVY
              PNRGIIAVALTRHHISIDAKSRVMEINMAIYNKRGKADLISHLSPFDSITKDL
              YLESTFVMEQDVLIEPFRSFKRPREKAVATVVLISLAETOKLKNVVLNIGSEPT
              IAVTFTMGVGYVMTCOMVSKSDVVSSESDNDELVEFSPADRCGLTTPCPNRY
              GNCIASCAKATKELVGRKGLVVAALGEALIERLHNEKVLADARTWLSSENGIP
              SKRFLGITGSPKFDYGVDFGWRPKAFDITSVDYAEILYVIOISDPFEGVIEGSLP
              KIHMAFAFKIIEFGCSLS"
              1679
              /note="24 a nucleotides"
BASE COUNT  488 a      350 c      359 g      482 t
ORIGIN
Query Match      98.6%; Score 1679; DB 7; Length 1679;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1679; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TCATTATGAGCAAAATCGTGAAGGTTCTTGAAAAATGCCAAGTTACACACCAT 60
    |||||||
Db 1 TCATTATGAGCAAAATCGTGAAGGTTCTTGAAAAATGCCAAGTTACACACCAT 60
Oy 61 CTGACACACAGATGTCGATTAATCGCTAACCAATTCCTGATATCCCTGGTTGC 120
    |||||||
Db 61 CTGACACACAGATGTCGATTAATCGCTAACCAATTCCTGATATCCCTGGTTGC 120
Oy 121 ACTTGAATTAAGATGACAGTCCCTCTGTTTACACATTTCCGTACCAAGAACCATTTCT 180
    |||||||
Db 121 ACTTGAATTAAGATGACAGTCCCTCTGTTTACACATTTCCGTACCAAGAACCATTTCT 180
Oy 181 TGACACACTGTTATCCCTTAATCTTAAGGCCCTTGTGTCCTACACTTAATAACACTAGCTTC 240
    |||||||
Db 181 TGACACACTGTTATCCCTTAATCTTAAGGCCCTTGTGTCCTACACTTAATAACACTAGCTTC 240
Oy 241 CGCTTAGCGGAATTTGTTGATCGCATCAAAATCGGCGGAATCGCGAAGTTTCAGTACT 300
    |||||||
Db 241 CGCTTAGCGGAATTTGTTGATCGCATCAAAATCGGCGGAATCGCGAAGTTTCAGTACT 300
Oy 301 CCCGTATAGGGCGACATGATTAATCTTGTGCTGCGAGTGTGACACAGATTTTGACT 360
    |||||||
Db 301 CCCGTATAGGGCGACATGATTAATCTTGTGCTGCGAGTGTGACACAGATTTTGACT 360
Oy 361 ACCTTAAGGTCATCAACGTGATTCCAATGATTTGCATGGCCCTTTTATGTTATGTC 420
    |||||||

```

Db	361	ACGTTAAAGGTCATCAACGCTGGTAGATTCCATCATATTTCGATGGCCCTTTTATGTTATG	420
Qy	421	CACGGGTTATTAGGACCATCATCAAGACTATTAAAGTAGATCCCGCTCGTAGCCGTGCAAGTAA	480
Db	421	CACGGGTTATTAGGACCATCATCAAGACTATTAAAGTAGATCCCGCTCGTAGCCGTGCAAGTAA	480
Qy	481	CCGTTTTTCCAAACCGGGGATAGCCGTGGCTGTGACGGGCAATCATTTCAATTGCAAGT	540
Db	481	CCGTTTTTCCAAACCGGGGATAGCCGTGGCTGTGACGGGCAATCATTTCAATTGCAAGT	540
Qy	541	CTAAAGATTGTTGTANTTTTCATCATATGCTTTGGCCATATTTAAACAATTTGGGAAAGACG	600
Db	541	CTAAAGATTGTTGTANTTTTCATCATATGCTTTGGCCATATTTAAACAATTTGGGAAAGACG	600
Qy	601	CGGACTTGTTGTCCGCGAATCTTTCATCTTTCGATAGATCGATAATCAAAAGATCTGT	660
Db	601	CGGACTTGTTGTCCGCGAATCTTTCATCTTTCGATAGATCGATAATCAAAAGATCTGT	660
Qy	661	ATGGCCATAGAGGAAACAATTTTGGACGAAATGCAATGTCTTGAAAGTCTCTAAT	720
Db	661	ATGGCCATAGAGGAAACAATTTTGGACGAAATGCAATGTCTTGAAAGTCTCTAAT	720
Qy	721	TTGGAGACCAACCCCTCGATTGCAAGGTAGAGCTACATATGCTCTCCCTTGCTG	780
Db	721	TTGGAGACCAACCCCTCGATTGCAAGGTAGAGCTACATATGCTCTCCCTTGCTG	780
Qy	781	AAATCTCAGAAAGCTTAAGACAAAGTACTGANTCTCAGAGATCCGAACCGACAATACGTG	840
Db	781	AAATCTCAGAAAGCTTAAGACAAAGTACTGANTCTCAGAGATCCGAACCGACAATACGTG	840
Qy	841	TAAAGAGGTTACAAATAGCTGTGGATAGCTATGAGCATATGATGGTCAATGCAAAAGATG	900
Db	841	TAAAGAGGTTACAAATAGCTGTGGATAGCTATGAGCATATGATGGTCAATGCAAAAGATG	900
Qy	901	ACGTCGTATCAGAGAAATTCATTCACAGACGACAAATAGCTCGAGTACTTCAGTTTACAG	960
Db	901	ACGTCGTATCAGAGAAATTCATTCACAGACGACAAATAGCTCGAGTACTTCAGTTTACAG	960
Qy	961	CGGATTGCCGAGACCTTCTAGCGCCCCCGCTGTCCGCTTAACACTTGTGGCAACGTCTGTG	1020
Db	961	CGGATTGCCGAGACCTTCTAGCGCCCCCGCTGTCCGCTTAACACTTGTGGCAACGTCTGTG	1020
Qy	1021	CGTCATCGCTTGCAAAAGCAACATTAAGAGTTAATTGGGGTAAAGGGCTTCTTGTG	1080
Db	1021	CGTCATCGCTTGCAAAAGCAACATTAAGAGTTAATTGGGGTAAAGGGCTTCTTGTG	1080
Qy	1081	CAGTTTGAGGCTATTGSGAGAGCCATTGAAAGAGTGTGACAAACGAAAGGCGTCTGTG	1140
Db	1081	CAGTTTGAGGCTATTGSGAGAGCCATTGAAAGAGTGTGACAAACGAAAGGCGTCTGTG	1140
Qy	1141	CAGATGCAAAAACCTTGGTTATCGGAATCTTATGGAATCCCTTCAAAAAGATTCTCGGGA	1200
Db	1141	CAGATGCAAAAACCTTGGTTATCGGAATCTTATGGAATCCCTTCAAAAAGATTCTCGGGA	1200
Qy	1201	TTACCGGATCGCTTAAGTTGATTCGTATGCTGTAGATTTTGGATGGGAAAGCCTGCA	1260
Db	1201	TTACCGGATCGCTTAAGTTGATTCGTATGCTGTAGATTTTGGATGGGAAAGCCTGCA	1260
Qy	1261	AATTTGACATTACCTCGTTGATTTATGCAAGAAATGATTTATGATTCAGTCCGCAAGGANT	1320
Db	1261	AATTTGACATTACCTCGTTGATTTATGCAAGAAATGATTTATGATTCAGTCCGCAAGGANT	1320
Qy	1321	TTGAAAGAAAGGTGTGAGATTTGGATATCATATGCTTAAGATTCAATAGCATGCAATTGGCA	1380
Db	1321	TTGAAAGAAAGGTGTGAGATTTGGATATCATATGCTTAAGATTCAATAGCATGCAATTGGCA	1380
Qy	1381	AAATCTTTGAAGAAGCTTTTGTCTTTGTGCATATGCTCTTTAATAGAACATATTTGCT	1440
Db	1381	AAATCTTTGAAGAAGCTTTTGTCTTTGTGCATATGCTCTTTAATAGAACATATTTGCT	1440
Qy	1441	GCAATTAAGATACCAAGTCTTTAGTAACTATACCAAAACCTACTTCTTGAGGGCGGAAAC	1500
Db	1441	GCAATTAAGATACCAAGTCTTTAGTAACTATACCAAAACCTACTTCTTGAGGGCGGAAAC	1500

QY	1501	ACCACAGCAGGTTCAATACACAGAAAGGTTGACTTCATAAATTCACAGAGTCGAATTATA	1560
Db	1501	ACCACAAACAGGTTCAATACACAGAAAGGTTGACTTCATAAATTCACAGAGTCGAATTATA	1560
QY	1561	CACCGTTGCTCTGTAAAGATTGAACCTTCACACCTGCACATGGTGTGTACATAGATTGT	1620
Db	1561	CACCGTTGCTCTGTAAAGATTGAACCTTCACACCTGCACATGGTGTGTACATAGATTGT	1620
QY	1621	AATAATGCCATTTATATACCTTCATTAAGTATCTATATGCATATAGAGAACAATGTTATGT	1679
Db	1621	AATAATGCCATTTATATACCTTCATTAAGTATCTATATGCATATAGAGAACAATGTTATGT	1679
RESULT	3		
LOCUS	E12757	1508 bp	DNA
DEFINITION	Senecio cruentus mRNA for acyltransferase,,partial cds.		24-JUN-1998
ACCESSION	E12757		
VERSION	E12757.1	GI:3251589	
KEYWORDS	JP 1997070290-A/5.		
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 1508)		
AUTHORS	Ashikari,T., Tanaka,Y., Fujiwara,H., Nakao,M., Fukui,Y., Yoshikura,K., Mizutani,M. and Kusumi,T., GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY Patent: JP 1997070290-A 5 18-MAR-1997;		
TITLE			
JOURNAL	SUNTORY LTD		
COMMENT	OS Senecio cruentus PN JP 1997070290-A/5 PD 18-MAR-1997 PF 30-JAN-1996 JP 1996046534 PR 17-FEB-1995 JP 95P 67159, 29-JUN-1995 JP 95P 196915 PI ASHIKARI TOSHIIKO, TANAKA YOSHIKAZU, FUJIWARA HIROYUKI, PI MASHIRO, PI FUKUI YUKO, YONEKURA KEIHO, MIZUTANI MASAKO, KUSUMI TAKAHI PI C12N15/09,A01H1/00,C07H21/04,C07K14/42,C12N9/10,(C12N9/10, PC C12R1:865), CC (C12N9/10,C12R1:19); PC strandedness: Double; CC topology: linear; FH key Location/Qualifiers FH source 1. 1508 FT /organism='Senecio cruentus' FT /tissue.type='petal' FT /clone='pCAT48', FT CDS 1.1367 FT /product='acyltransferase'. FEATURES source Location/Qualifiers 1. 1508 /organism='unidentified' /db_xref='taxon:32644' BASE COUNT 442 a 293 c 296 g 477 t ORIGIN		
Query Match	12.0%:	Score 204;	DB 5; Length 1508;
Best Local Similarity	52.1%:	Pred. No. 3.2e-41;	
Matches 721; Conservative	0;	Mismatches 615;	Indels 48; Gaps 10;
QY	25	TGAAGTCTTGAAGAAAGCCAGTTACACCCACCCACCTGCATGACACACAGATGTCAGATTAT	84
Db	1	TGAACATTTCTCAACATGCGCCGAAATATGCGCCCCCTCGGGACCA--TCGGCCATCGCT	57
QY	85	CGCTACCGGTACATTTCTTCGATATCCCTCGGTGCACATTGAATGAAGATGCAGTCCCTTC	144
Db	58	CGTTATCTCTTACTTCTTCGACATTACTTGGCTACTCTCCCTCGGTCCACACATCTTT	117
QY	145	TGTATTAGACTTTCCGTACCCAGAAACACATTTCTTGACACGTGTTATCCCTAATCTTA	204

Db	118	TCCTCATAGACTTTCACAACTTCTAAATCCAAATTTATGAGACATATTGTTCCAGGCTAA	177
Qy	205	AGGCGCTTTTGTCTCTCACTCTTAAACACATCACTGTTCCGCTTAGCGGAATTTGTTGATGC	264
Db	178	AACAACTTTATCGGCACTCTTCAACATTTTTCCTCCGTTTGATGAAATTTGATTTGAT	237
Qy	265	CGATCA-----ATGGGCGAAATGCCGAATTTCACTACTCCCGATGAGGCG	315
Db	238	TTCTTAACACTGATGGTTTCGGGTTTTTAATTAATAAACCAAAATTAACACGTTGAAGTG	297
Qy	316	ACTCGAATCACTTGTGATCGTTGGCGAGTCCGACAGATTTTGTAGACTTAAAGGTATC	375
Db	298	ATTCTGTGTGTTACTTTTGAGAAATGTGTCTGACTTAATTAATTTGACGAAATC	357
Qy	376	AACGTGTATATCCCAATGATTTGCAATGCCCTTTTATGTTATGCCAGGGTTATAAGA	435
Db	358	ATCTCTGAAAAATGTAATAACTTTTATCCACTGTACCTGCATG---GGAATATGCATCA	414
Qy	436	CCATCGAAGACTATTAAGATATCCCGCTCGTAGCCGCGAAGTAACGCTTTTCCATAC	495
Db	415	AAATATATGATTTGGCTCAAGGCTCCACCTTTTCTACTTCAAGTACCTTTTTCGGGCT	474
Qy	496	GTGGCATACCGCTGGCTGTAGCGCACTATTCATATTCAGATGCTAAAGTTTGTAA	555
Db	475	CGGGATATACATAGAAATGACGATCATCATCTTACCTTGTGACGTACGACCGGTCA	534
Qy	556	TGTTATCAATCTTTGGGCGCTATATTAACAAATTTGGGAAAGACGCGACTTGTCCG	615
Db	535	ACTTTTGAAGGGTGGCACTTCGATTAATCAATCGGTATGATCGCTCTTTTATACGA	594
Qy	616	CGAATCTTCTTCATCTTTCGATATGATCGATATCAAAAGATCTGTATGGCCTAGAGAA	675
Db	595	AAGGATCTCCACCGGTTTTTATAGATGATTACA-----TCCACATTTAGATGA	646
Qy	676	CATTTTGAAGAAATGCAAGATGTCTGTGAATGCTTCATATTTGGAAGAACCC	735
Db	647	AAATAGTTTGAACATACAG-----GCTCGAAGTTTTTATTAACCTGTGAGCTTGTG	702
Qy	736	CTCGATTCACAGATAGCACTCATATGTCTCTCCCTTGGCTGAATACGAAGCTAA	795
Db	703	GTCCCACTGAATAAAGTTGCGTCAACGTTGTGTTGACCGCACTAATATCAATCTACTA	762
Qy	796	AGAACTGCTACTGATCTCAGAGATCCGACCCGACAAATACGTATACGACGTTCAAA	855
Db	763	AGAAAAAGCTTAA-----CCAAAGTCCAAACTGTGAGACATGT--CATCTTTACGG	816
Qy	856	TGACGTGTGATAGTATGACATGCATGTGTCAAATCAAAAGATGACGTCGATACAGG	915
Db	817	TAACTTGGTTATATATATGAGGTTGCATACGAAATCA-----CTCGTAAAAATAG	867
Qy	916	AATCATGCAACGAGAAATAGCTCGATGACTTCAAGTTTACAGGGATTCGAGAGAC	975
Db	868	GAGAAAGAAAGGGGAGAGACGATAGAACAGTTCAATACCATTTGATTCGATCTC	927
Qy	976	TTTTCAGCCCCCTGTGCCCTTACACTTGGCAACGTGTGGCTATACGTTGCA	1033
Db	928	GTTCTGATCCACAAATTCACACAGCCTACTTGTGTACTGTGTGACACCATGTGTCCGA	987
Qy	1036	AAGCAACACATAAAGATTTGTTGGGATTAAGGCGCTTCTGTGCAGTTTCAGCTATTG	1095
Db	988	CTTTAAAAAATGTCGTTTACTAGGAAATAGGTATACCTTGGTGTCTAAAGTAATG	1047
Qy	1096	GAGAAGCCATTGAAAGAGGTGTCACACAGCAAAAGAGCGTTCGTGAGATGCAAAACTT	1155
Db	1048	GAGAGTCTATATGCAAAAT	1107
Qy	1156	GGTATTCGGAATCTAATGAATCCCTTCAAAAAGATTTTCGGGATTTACCGGATGCCCTA	1215
Db	1108	GG---CATGAACCTTTCATGATCCCGGCTAGAAAGAT---GGTGTGCTGTACACTA	1161
Qy	1216	AGTTGAGTTGTAATGAGTGTGATTTTGGATGGGAAAGCTCCAAATTTGACATTA	1275
Db	1162	AGCTCACTGTAGACTTGTGATTTTGGGTGGGGAAGCGGTAAGATATGAGACTGTTT	1221

QY	1276	CTGTGATTTATGCAAGAAATTTATTTATGTGATTCAGTCCAGGATTTTGGAAAAAGGTGCG	1335
Db	1222	CAATGACCTTAATTAACCTCATCTTCTATTAATGCAAGCAAAACATCAGACAGATCTTG	1281
QY	1336	AGATTGGATGATCATGCGCTTAAGATTCATATGATGATTCGATTTGCAAAAATCTTGAAGAAC	1395
Db	1282	AAATGATGATGATGATTCACGAGATATGCAAAATGAGGCGTTCTTACATCTTTGATGAAG	1341
QY	1396	GCTT 1399	
Db	1342	GATT 1345	
RESULT	4		
LOCUS	E12754	1622 bp	DNA PAT 24-JUN-1998
DEFINITION	Gentiana triflora mRNA for acyltransferase, complete cds.		
ACCESSION	E12754		
VERSION	E12754.1	GI:3251586	
KEYWORDS	JP 1997070290-A/2.		
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 1622)		
AUTHORS	Asahikari, T., Tanaka, Y., Fujiwara, H., Nakao, M., Fukui, Y., Yoshikura, K., Mizutani, M. and Kusumi, T.		
TITLE	GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY		
JOURNAL	Patent: JP 1997070290-A 2 18-MAR-1997;		
COMMENT	SUNTORY LTD		
OS	Gentiana triflora		
PN	JP 1997070290-A/2		
PD	18-MAR-1997		
PF	30-JAN-1996 JP 1996046534		
PR	17-FEB-1995 JP 95P 67159, 29-JUN-1995 JP 95P 196915		
PI	ASHIKARI TOSHIMIKO, TANAKA YOSHIKAZU, FUJIWARA HIROYUKI, PI NAKAO MASHIRO,		
PI	FUKUI YUKIO, YONEKURA KEIKO, MIZUTANI MASAKO, KUSUMI TAKAAKI		
CI	C12N15/09, A01H1/00, C07H21/04, C07K14/42, C12N9/10, (C12N9/10, PC		
CI	C12R1:865)		
PC	(C12N9/10, C12R1:19);		
CC	Strandedness: Double;		
CC	Topology: Linear;		
FH	key	Location/Qualifiers	
FH	source	1.1622	
FT	source	1.1622	
FT	/variety='japonica'	/organism='gentiana triflora'	FT
FT	/issue_type='petal'		
FT	/clone='pgAT106'		
FT	CDS	35..1474	
FT	Location/Qualifiers	/product='acyltransferase'.	
FEATURES	source	1.1622	
BASE COUNT	478 a 345 c 322 g 477 t		
ORIGIN			
Query Match	11.14; Score 189.2; DB 5; Length 1622;		
Best Local Similarity	52.2%; Pred. No. 1.9e-37;		
Matches 728; Conservative	0; Mismatches 618; Indels 48; Gaps 12;		
QY	25	TGAAGGCTCTTGAAATAAGCCAGATTATACCAACCATCTGCACACAAACAGATGCGAGTTAT	84
Db	57	TCAAAGTCTTGAGAAATGCCGTGTGGCCACACACCGGAC--GCCGTCCCGAGTTTA	113
QY	85	CGCATCGGTAACAATCTTCATATCCCTGCTGCACCTTGAATTAAGATGCGACCTTC	144
Db	114	CAGTCCACCTGCTGTTTTCACATGCGATGTGTTGATCTCTGATGCGAACAACCATCTGC	173
QY	145	TGTTTTAGACATTTCCGTACCAAGAACACATTTCTTGAGACATGTTATCCCTAATCTTA	204

```

Db 174 ATTTCAGAGATCCGCCATCCCTTGCCAA---CTCTAAATTTATCATTTATCCATTA 230
Oy 205 AGGCGCTTTGTCCTCACTATAAACACTAGCTTCCGGTATGCGGAATTTGTTGAC 264
Db 231 AATCGCCCTTCCCTGCTGCTCAAAACACTTCTTCCGTAGCCGGGAATTTGATTTGCC 290
Oy 265 CGATCAAAATGCGGCGA---AATGCCAAGTTTCACTCCCGTGATGAGCGACTGCA 321
Db 291 CGGTAGATTCCTCGATAGAAATGCCGAGTGCCTTAC-----AAGAAGGGGACTCCG 344
Oy 322 TTAAGTTGATTCCTGCGGAGTGTGACAGAGATTTTACATCTTAAAGTCAATCACTG 381
Db 345 TTTCTTAAACAATTCGAGATGAGATTTGATTTATCTCGCCGAGATCATCAGA 404
Oy 382 TAGATTCCAAATGATTTGATGCGCTTTTATGTTATGTCAGCGGGTATTAAGAACATGC 441
Db 405 GGGATTTCTTATTAATTCAAACGATTTGATTTCCGACGTGCCAGAACCGATTTGATACC---T 461
Oy 442 AAGACTATTAAGTATCCCGCTCGTAGCCGTGCAAGTAACCGTTTTCCTAACCGTGCA 501
Db 462 CCGGGAGACAAGTATTTACACTTTTTCCTTACAGTGACGGTGTCTTCCAAACCGGTA 521
Oy 502 TACCGCTGCTTGACGGCACATCATTTGACAGATGCTAAAGTTTGTATGTTCA 561
Db 522 TATGATTTGGACGCAATCTTCAAGTTCTTGATGATCCAGTTCTTTCTGCAATTTA 581
Oy 562 TCAATGCTTGGCTATATTACAAAT---TTGGGAANAACCGGCACTGTTGTCCGCA 618
Db 582 AATAATTTATGGTTTGGTTGTAACAAATCCAAAGAGATTCATTAAGTCTTCCACTTT 641
Oy 619 ATCTTCTTCATCTTTCGATAGATGATTAATCAAGATCTGTATGGCTAGAGAAACAT 678
Db 642 CTTCCTACCTATGATAGACACATGCTGTGTGCAAGATTCATTTATCTCGAAAA 701
Oy 679 TTTGGAACAATGCAAGATGT---TCTTGAATGTTCTTGAATTTGGAACAAACCC 735
Db 702 TCTACATGTAAGAAACACTGCTCAATCTCAGGCAACACTGCTTCAATATCCAGCA 761
Oy 736 CTGATTCMAAGTAGACGATACATATGCTCTCCCTGCTGAAATCCAGAACTAA 795
Db 762 TTTCTAAGATGAATGTCAGACACCTTCTATCTTACACCTTATGATATCAAGAGTCA 821
Oy 796 AG-AACAAAGTACTGAATCTCAGAGATCCGAAACGCAATA-----CGTGAACGA 846
Db 822 ACAAATTCATTTGCTCAAAAAAATCGCACTTAACCGGTAGTATATATATCTGTCA 881
Oy 847 CGTTCAATGACGTGTGATACGTATGACATGCTGCTCAATCAAAATGACGTGC 906
Db 882 CTTCACGGTACATCTGACATGATCTGACATGCTTGTGAAATCATTTAGACACCGTGC 941
Oy 907 TATCAGAGAAATCATGAAAGGAAATGAGCTCGAGTCAAGTTTACAGCGGAT 966
Db 942 TAGAGAGAAA---GGTGAAGAGATTAACATGCAACAACTTATGTGCTTCACTCACT 998
Oy 967 GCCGAGACTTGTGACGCCCGCTGCTGCTCACTACTTGTGCAACTGTTGCGCTCAT 1026
Db 999 GCCGACACGTTTGTGCTGCGGATACCTCAAAATTTACTTGGAAATGTCATAGTGCCT 1058
Oy 1027 GCGTTGCAAAACACACATTAAGATTAGTTGGGATTAAGGCTTCTTGTGCACTTG 1086
Db 1059 GTATGTGGGATGCACTCTGAGCAACTTGTAGAAATGAAGGTTGTGCGTAGCTGCA 1118
Oy 1087 CAGCTATTGGAAGACATGAAAGAGGTGCAACAGAAAAAGCGTTTTCAGAGT 1146
Db 1119 CCGGCATCGGAGATGCTATCCATTAAGAGTTACATGACATGCAAGAAATTTCTGAGAG 1178
Oy 1147 CAAAAAATGTTATCGAATCTA-----ATGGAATCCCTTCAAAAAATTTCTCG 1197
Db 1179 ATTGATATATCCGCCCGCATACATCTGGGCAACAAAGTGCAGCGCTATTTATGTGC 1238
Oy 1198 GGATTACCGGATCGCCTAAAGTTGATGTTGTTAGATTTTGGATGGGAAAAAGCTTG 1257

```

```

Db 1239 TTGATCCGACACAGCAATGTGCATGATTTGATGACAGATTTGGTGGGAAAGCTTG 1298
Oy 1258 CAAATTTGACA---TTACTCTGTTGATTATGCAAGATTTGATTTGATTTGATGCTCA 1314
Db 1299 AAAAGCATATCTGTTTAACATCTCTTGGCAACACTAAATTTTATGATCTCGGTCA 1358
Oy 1315 GGGATTTGAAAAAGGTGTGAGATTGAGTATCATTTGCTTAAGATTTATGATGAT 1374
Db 1359 GAAGATTTAAGAGACACTTGAGCTTGCAATTTCTTGGCTAAGAAATAGATGAGCGAT 1418
Oy 1375 TTGCAAAAATCTTT 1388
Db 1419 TTGCCACCATTTT 1432

```

```

RESULT 5
AB026494
LOCUS AB026494 1622 bp mRNA PLN 04-APR-2000
DEFINITION Gentiana triflora GAT106 mRNA for acyltransferase homolog, complete cds.
ACCESSION AB026494
VERSION AB026494.1 GI:7415596
KEYWORDS acyltransferase homolog.
SOURCE Gentiana triflora CDNA to mRNA.
ORGANISM Gentiana triflora
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Asteridae; euasterids I; Gentianales; Gentianaceae; Gentiana.
REFERENCE 1 (sites)
Yonekura-Sakakibara,K., Tanaka,Y., Fukuchi-Mizutani,M., Fujiwara,H., Fukui,Y., Toshihiko,A., Yanaguchi,M. and Kusumi,T. Molecular cloning and biochemical characterization of hydroxycinnamoyl-CoA:anthocyanin 3-O-glucoside 6-O-hydroxycinnamoyltransferase from Perilla frutescens and diverse plant acyltransferase homologs Unpublished (1999)
2 (bases 1 to 1622)
AUTHORS Tanaka,Y. and Yonekura-Sakakibara,K.
TITLE Direct Submision
JOURNAL Submitted (21-APR-1999) to the DDBJ/EMBL/GenBank databases.
Yoshikazu Tanaka, Sunitory Ltd., Institute for Fundamental Research; Wakayama-dai 1-1-1, Shimanoto, Osaka 618-8503, Japan
(E-mail:Yoshikazu.Tanaka@sunitory.co.jp, Tel:81-75-962-8807, Fax:81-75-962-8262)
FEATURES
source
location/Qualifiers
1..1622
/organism="Gentiana triflora"
/db_xref="taxon:55190"
35..1474
/gene="GAT106"
35..1474
/gene="GAT106"
/codon_start=1
/product="acyltransferase homolog"
/protein_id="BAA93452.1"
/db_xref="GI:7415597"
/translacion="MAGSEDIKYLEKCRVAPPDVAEFTVPLSPFDMRWLISDAEH
HLHFRPBPCKNSKFTISSTKSLSEKLEKPLAGLILPVPSSRMPELRYKFSN
VSUETIASSMDPTLADDRORDSKFNDLIPQLPPLVTSGDVLPALQVTFNS
TGICIGRNLIHOVLDDASSFLHFNKLWLVDSNSDSLKFLPLSSLPMDRVSVDPPH
IRRKIYNERKLKLSOGTPTVLPNPAISKDEVAFTLHPIDIMLKLFKFSKRNRLTGS
SNVNSTPTVTSALIMTCLSKSLDVTVEKEEDKHAANLCAFINCRORFAPIPQNY
FGNCIVPCMGSTHBOULGNEGSLVAATADLAIHKRLHOVEGLIRGDMWISPPYSTA
APRSGLIYVGSQARNYHDPDADCGMKLEKHESVSTNPSTLILISRSRFRKALEL
GISLPKRMDFATITFTNLSLHVSPL"
BASE COUNT 478 a 345 c 322 g 477 t
ORIGIN

```

```

Query Match 11.1%; Score 189.2; DB 7; Length 1622;
Best Local Similarity 52.2%; Pred. No. 1.9e-37;
Matches 728; Conservative 0; Mismatches 618; Indels 48; Gaps 12;

```

QY 25 TGAAGTCTTGGAAAAATGCCAACTTACACACCATCTGACACAGATGTGCATTTAT 84
 DB 57 TCAAAAGTTCTTGAGAAATTCGCCGTGTTCGCCACACCGSAC---GCCGTGCCGAGTTTA 113
 QY 85 CGCTACCGGTAACTTCTTGATATCCCTGGTGTGACACTTGAATTAAGATGACAGTCCCTTC 144
 DB 114 CAGTCCACACGTGTGTGTGTGACATGCGATGTTGATCTGTGATGCAAGAACACCATCTGC 173
 QY 145 TGTATTACGATTCCTGACCAAGAACACATTTCTGACACTGTTATCCCTAATCTTA 204
 DB 174 ATTCTACAGATTCGCCGATCTGTGCCAA---CTCTAATTTATCATTCATTCATTGA 230
 QY 205 AGCCCTCTTGTCTCACTCTAAACACTACGTTCCGTTTACGGGAAATTTGTTGATCC 264
 DB 231 AATGCTCCCTTTCCTTGTCTCAACACTTTTCTGCTTACGCCGGGAAATTTGATTTGGC 290
 QY 265 CGATCAAAATCGGGCGA---AATGCCAAGTTTCAGTACGCCCGGTGATGAGGGCGACATCGA 321
 DB 291 CGGTAGATTCCTCGATAGAAATCCCGAGTGTGGTTAC-----AAGAAAGGGAGCTCCG 344
 QY 322 TAACATTGATCGTTGGGAGTCTGACCAAGATTTGATGACATTAAGCTTAAAGTCAATCTG 381
 DB 345 TTTCTTTAAACATTTGAGATGAGATGAGATGATTTGATTTATCTCGCCGAGATCATCAGA 404
 QY 382 TAGATTCGAATGATTTGATGAGGCTTTTATGTTATGCCACGGGTTTAAAGACCATGC 441
 DB 405 GGGATCTTAAATTTCAACGATTTGATTCGACAGCTGCCAACAACCGATTTGATACC---T 461
 QY 442 AAGACTATTAAGTATCCCGCTCGTAGCCGTCAAGTAAACCTTTTTCCTTAACCGTGC 501
 DB 462 CCGGGACAGAAATTTACCACTTTTGTGCTTACAGGTGCGGTGTTCTCAACACACCGGTA 521
 QY 502 TAGCCGTGCTCTGACGGCAGATCATTAATGATGATGCTTAAAGTTTGTGAATGTCA 561
 DB 522 TAGCATTTGAGCAGATCTTCATCAAGTTCTTGGTGTGATGCCAGTTCTTTTCGATTTTA 581
 QY 562 TCAATGCTTGGGCTATTTTAAACAAT---TTGGGAAAGACGGGACTTGTGTCCGCGA 618
 DB 582 ATAAATTAATGTTTGGTTGTGACAAATCCAAATGGAATTCATTAAGTTCTTCCATCTT 641
 QY 619 ATCTTCTTCATCTTTCATGATGATGATATCAAAAGATCTGTATGCGCTTAAGAGAAACAT 678
 DB 642 CTTCCTACATATGACAGATGATGTGTGACAGATCAATTCATATTCGTGGAATA 701
 QY 679 TTTGGAACGAATGCAAGATGT---TCTTGAATTTCTCTAGATTTGGAAGCAAAACCC 735
 DB 702 TCTACAAATGAAGAAACTGCTCAAAATCTCAGGGCAACCTAGTCTTCTAATCCAGCA 761
 QY 736 CTCGATTCAACAGGTAGAGCTACATATGCTCTCCCTTGTGAAATCCAGAGCTAA 795
 DB 762 TTTCTAAAGATGAAGTTGAGCCACCTTCATCTACACCTTATGATATGAGACCTCA 821
 QY 796 AG-AACAAAGTACGAATCTCAGAGATCCGACCGACACAATA-----CGTGTACGA 846
 DB 822 AGAATTCATTTCTCAAAAAAATGCAACTTAACCGGTATGTAATTAATTAATCTGTCAA 881
 QY 847 CGTTCACATGACCTGTGAGATACGTATGACATGATGCTCAATCAAAAGATGACCTGC 906
 DB 882 CTTCACAGGTGACATGCTGACTGATCTGACATGCTTGTGAATCATTTAGACACGCTGC 941
 QY 907 TATCAGAGGATCATGCAAGAGCAAAATGAGCTGAGTACTTCACTTTTACAGCGGAT 966
 DB 942 TAGAGAGAA---CGTGAAGAGAGATTAACATGACAGCAAACTTATGTGCTTCATCACT 998
 QY 967 GCCGAGGACTTGTAGACCCCGCTGCTCCGCTACTACTTGTGCAACTGTCTTGCGCAT 1026
 DB 999 GCCGACAACTTTTGTGCCCGCATACCTCAAAATTAATTTGGAATTTGATATGTCTTT 1058
 QY 1027 GCGTTGCAAAAGCAACAAGAGATTTAGTTGGGATTAAGGCTTCTTGTTCAGATTG 1086
 DB 1059 GTATGTTGGGATGACCTCATGACCACTTGTAGAAATGAGGGTTGTGGTAGCTGCA 1118
 QY 1087 CAGCTATTGGAGAGCAATTTGAAAAAGGTTGCAACAAGAAAAAGCGTTCTTTCAGATG 1146

DB 1119 CCGCCATTCGAGATGCTATCATTAAGAGTTTACATGACTACAGAGAAATTTGAGAGAG 1178
 QY 1147 CAAAACCTTGGTTATTCGAATCTA-----ATGGAATCCCTTCAAAAAAGATTTCTTC 1197
 DB 1179 ATTGGATTATCGCCGCCGCCGATCAACATCTGCGGCAACAAAGTGCAGCGCTATTATGTCG 1238
 QY 1198 GGAATTACCGGATCGGCTAGTTCGATTCGATAGTGTGATTTTGGATGGGAAAGCCCTG 1257
 DB 1239 TTGATCCGCAACAGCAATGTGATGATTTTGAATGACAGATTTGTTGGGAAAGCTTGG 1298
 QY 1258 CAAAATTTGACA---TTCCTCTGTTGATTTATGCAAAATTTGATTTATGATTCAGTCCA 1314
 DB 1299 AAAACATGAATCTGTTCAACTATTCCTTGGGCAACATATTTTGTATCTCTGCTGCTCA 1358
 QY 1315 GGGATTTGAAAAAGGTGTGAGATGAGATGATTCATTTCCCTTAAGATTCATATGATGCAT 1374
 DB 1359 GAAGATTTAAAGAGACCACTTGAGCTTGCAATTTCTTCTTCCTAAGAAATGATGACGCAT 1418
 QY 1375 TTGCAAAAATCTTT 1388
 DB 1419 TTGCCACCATTTT 1432

RESULT 6
 E12756
 LOCUS
 DEFINITION
 E12756
 ACCESSION
 VERSION
 E12756.1 GI:3251588
 KEYWORDS
 JP 1997070290-A/4.
 SOURCE
 ORGANISM
 unidentified.
 unclassified.
 1 (bases 1 to 1479)
 REFERENCE
 Ashikari,T., Tanaka,Y., Fujiiwara,H., Nakao,M., Fukui,Y.,
 Yonekura,K., Mizutani,M. and Kusumi,T.
 GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY
 Patent: JP 1997070290-A 4 18-MAR-1997;
 JOURNAL
 SUNTORY LTD
 OS Perilla oclimoides
 PN JP 1997070290-A/4
 PD 18-MAR-1997
 BP 30-JAN-1996 JP 1996046534
 PR 17-FEB-1995 JP 95P 67159, 29-JUN-1995 JP 95P 196915 PI
 ASHIKARI TOSHIMIKO, TANAKA YOSHIKAZU, FUJIMURA HIROYUKI, PI NAKAO
 MASAHIRO,
 PI FUKUI YUKIO, YONEKURA KEIKO, MIZUTANI MASAKO, KUSUMI TANAOKI PI
 C12N15/09,A01H1/00,C07H21/04,C07K14/42,C12N9/10,C12N9/10, PC
 C12R1:865),
 PC (C12N9/10,C12R1:19);
 CC strandedness: Double;
 CC topology: Linear;
 FH Key
 FH Location/Qualifiers
 FT source 1..1479
 FT /tissue_type='leaves'
 FT /clone='psat208'
 FT CDS 1..1343
 FT /product='acyltransferase'.
 FT location/Qualifiers
 1..1479
 /organism='unidentified'
 /db_xref='taxon:32644'
 BASE COUNT 420 a 316 c 331 g 411 t 1 others
 ORIGIN

Query Match 10.1%; Score 172.2; DB 5; Length 1479;
 Best Local Similarity 49.5%; Pred. No. 3,86-33;
 Matches 701; Conservative 0; Mismatches 668; Indels 48; Gaps 8;

78 GAGTATCGCTACCGGTAACATTTGATATCCCTGGTGCATTTGAATAGATGAC 137
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 48 GAGCAATCGGTCGCTACATTTCTGACATGATGCTGCTGATTTTCATCCATGCTT 107
 138 TCCCTCTCTTTAGCATCTTCCGTAACCAAGACATTTCTTGACACTGTTATCCT 197
 108 CAGCTCCCTCTTACGAATTCCTGTTCCAAAGCAACATTTTTCAGAAATCCATCGTTCCA 167
 198 AATCTTAAGCCCTCTTCTCTCTCACTTAAGAACTAGTTCCTGCTTGGGAAATTTTG 257
 168 AACTCAACAACTCTCTCTTAACCTCTCATACACTTTCTCTCTCTCATGCAATTTTA 227
 258 TTGATGCCGATCAATCGGGGCAATGCGAAGTTTCAGTACTCCCGTATGAGGCGAC 317
 228 ATCTACCCCTTCAATCCCGGAGAAATGCGGAGTTTCGGTA-----TCTATCCGGGAG 281
 318 TCGATTAACCTTATGCTTCCGAGCTGTGACAGAGATTTTGACTACTCTTAAAGTCA 377
 282 TCGGTTTCTTTCACATCCGAAATCTAGCAGACTTGTGATCTCGTGGAAATTCGT 341
 378 CTGTGATTTCCAAATGATTTGATGCGCTTTTATGTTATGTCACAGGGGTATTAAGACC 437
 342 CCAGATCTCCCGTTAGGCTCTACAACTTTGTCCCTAAATTTGCCGCCATTTGCGAAGA 401
 438 ATGCAAGACTATTAAGTATCCCGCTGTAGCCGTGCAAGTAACGCTTTTCTTAACGCT 497
 402 ---TCCGATAGAAACTCTTCCAAAGTTTTCGCGTGCAGTACTCTTCCAGAGCCGA 458
 498 GGCATAGCCGTGGCTGTGACGCGACATCTCAATTCGATTCGATGCTTAAAGTTTGTAAAG 557
 458 GGCCTGCTGATTTGATAGATAGCAACGCTACACGCTTGAAGCGGCCCGCTGTTCTCGCG 518
 558 TTTCATCAATGCTTTGGCCCTATTAACAANA---TTTGGGAAAGACGCGAGCTTGTGTC 614
 519 TTATTAACGCGCTTGGCTTCAATGAGCAACAACATTTGAATGAATGAATGAAGA 578
 615 GCGAATCTTCTTCATCTTTGATAGATGATCAATCAAAATCTGTATGCGCTAGAGAA 674
 579 TTATTAATCTTCTGATTTTTCATGATCCGATCAATTAATTAATCCGCAAAATTTGCTCC 638
 675 ACATTTTGAACGAATGCAAGATGCTCTTGAATGTTCTGATGATTTTGAAGCAAAACC 734
 639 ATTTATTTGAAGAACGCGCTAAATTTCT-----TTGCATCTGCTATCCCTTA 689
 735 CCTGATTCACAAAGTATGAGCTATCATATGCTCTCCCTTCTGTAATCCAGAACTA 794
 690 TTACGAGCGAGCGGATTCGAAACAGCTTCTGTTTACCCATTCCAAATTAAGAAATG 749
 795 AAGAACAAGTACTGATCTCAGAGATCCGAACGCAATAGCTGTAAAGACGTTTACA 854
 750 AAG-----GTTTGAATTCAGTCCAGAGTCCAAAGTTTGTAGTCCATCTCATCTTTTGT 803
 855 ATGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 914
 804 GCGATTTGACGCTTATATGTTGGCTGCGATTAACGAATTAATTAACAGCAATGAAG----- 858
 915 GAATCATGACGAGCAAAATGAGCTCGAGTACTCATGTTTTCACAGCGGATTTGCCAGGA 974
 859 -----ACCAAGACAAACGAGATGATTTTCTTCTGATTCGCGTGCATCTAAGGCCA 908
 975 CTTCGAGCGCCCGCTGTCCGCTTAATCTTGTGCACTGTCTTCCGTCATGCGTTGCA 1034
 909 CATATTAAGATCCCGGCTTCTCAAAATTAATCTTCCGGAATGCTTATCTACACGCGTCCG 968
 1035 AAGACACATTAAGAGTTAGTTGGGATTAAGGCGCTCTTGTGCACTTGCATGCTATTT 1094
 969 ACAAATGCCGCGCGAGAGTGTGGAGAGAAAGGGGTCTTCTGCGACACTGAGGTAATC 1028
 1095 GGAGAGACCATTTGAAGAGTTTGCAACAAGAAAGGCTTCTTGCAATGCAAAACT 1154
 1029 GCGGCGGAGATTAATAAAGAT---CAACGACAAAGAAATTTAGAAACGCTGAGGAA 1085
 1155 TCGTATGCGAATCTAATGGAATCCCTTCAAAAAGATTTCTCGGATTAACGGAATCGCT 1214

111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 1086 TGTGCGCCGAGATGCTGAAGGTTGACAGAAATCATATTTTTCGCTGCGAGATGAC 1145
 1215 AAGTCGATTCGTATGCTGATGTTTGGATGGGAAACCTGCAAAATTTACATTTACC 1274
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 1146 AAGCTGATCTTTACGCTGACATTTTGGATGGGAGCGCAAGCAACAAATATTTG 1205
 1275 TCTGTGATTAAGCAATTTG---ATTATGATTCACCTCCAGGATTTGAAAGAT 1331
 1206 TCGATTTAGGGGAGAAATATGATGACCTTTGTAAGCCAGGGATTTGACAGAGGA 1265
 1332 GTGAGATTTGAGATATCATTTGCTTAAGATTCATGATGATTCATTTGCAAAATCTTTGAA 1391
 1266 TTGAGAGCTTGTCTGCTTGTGCTTAAGACAAATGATGCTTTGCTGCTTATTTTCA 1325
 1392 GAAGCTTTTGTCTCTTGTGATAGTCTCTTTAATGAAACCATTTTGTCTGCAATTAAGTA 1451
 1326 CTGGGAATTAATGTTAAATTAATGTAATTAATTAATTAATTAATTAATTAATTAATTA 1385
 1452 CCAAGCTTTTATGATTAACACTACACCAACCTACTTT 1488
 Db 1386 TTAAGTGTGATTAAGCTAAGCAATTAATCCCTATTTAT 1422

RESULT 7
 AB029340 1476 bp mRNA PLN 04-APR-2000
 LOCUS
 DEFINITION Perilla frutescens mRNA for anthocyanin acyltransferase, partial cds
 ACCESSION AB029340.1 GI:7415645
 VERSION
 KEYWORDS anthocyanin acyltransferase.
 SOURCE Perilla frutescens leaf cDNA to mRNA.
 ORGANISM Perilla frutescens
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Lamiales; Lamiales; Perilla.

REFERENCE 1 (sites)
 AUTHORS Sakakibara, K. Y., Tanaka, Y., Mizutani, M. F., Fujiwara, H., Fukui, Y., Ashikari, T., Yamaguchi, M. and Kusumi, T.
 TITLE Molecular and biochemical characterization of a novel hydroxycinnamoyl-CoA: anthocyanin 3-O-glucoside-6 unpublised (1999)
 JOURNAL 2 (bases 1 to 1476)
 REFERENCE Unpublished (1999)
 AUTHORS Sakakibara, K. Y. and Tanaka, Y.
 TITLE Direct Submission
 JOURNAL Submitted (24-JUN-1999) to the DDBJ/EMBL/GenBank databases. Keiko Y Sakakibara, Suntory Research Center, Fundamental Research, Plant Biotech, Shimamoto-cho, Wakayamada, 1-1-1, Mishima-gun, Osaka 618-8503, Japan (E-mail: Keiko_Sakakibara@suntory.co.jp, Tel: +81-75-962-8807, Fax: +81-75-962-8262)

FEATURES

source

1..1476

/organism="Perilla frutescens"

/db_xref="taxon:48386"

/tisue_type="leaf"

<1..1343

/product="anthocyanin acyltransferase"

/protein_18="BA93475.1"

/db_xref="GI:7415646"

/translation="VTEICRVGPPDPSVAEQSVLPTEFDMTLMHFHMLQLPYEPFC SKOHSESIIVKLIKOSLSKTLHFEPISCNLIYSPSPKMEPFLYLSQDSVFTIAES SDPDDIVGNRPESPVRLYNFPLPTEVSDRLQVFAVOYTLFPGRCVGIAT HHYSDASFLAFTIAMSMSKHLENDEDEPFLPDPDSVYIKYPRKPSYWRNA LKFLQSRHSLPFDRIKTTFTVFSQSKTKLKGWIOSVPSLVHLSFVATAAYMMAG ITKSEPTADEDDNDADAFLLPVDLRPLRDPVPENYGNCSYVALPRMRRELVEGK VFLAEVIAAEIKRKIRINDKRLIEVEKSKYIRKALQKSYFVAGSSKLDLYGDFGW GKARQETLSDIGEVYAMTLCARDPEGLVCISLPKDKMDAAVAFFSLGIC"

BASE COUNT 419 a 315 c 331 g 411 t

ORIGIN

Query Match 10.1%; Score 171.4; DB 7; Length 1476;
 Best Local Similarity 49.9%; Pred. No. 6,1e-33;
 Matches 707; Conservative 0; Mismatches 661; Indels 49; Gaps 9;

78 GAGTATACGCTACCGGTAACATTCCTTCGATATCCCTGCTTGACACTTGAATAGATGAC 137
 48 GAGCAATCGGTCGCGCACATTCCTTCGATATCCCTGCTTGACACTTGAATAGATGAC 107
 138 TCCCTTCGCTTACGACTTCCTTCGATATCCCTGCTTGACACTTGAATAGATGAC 197
 108 CAGCTCTCTTCACGATATCCCTGCTTGACACTTGAATAGATGAC 167
 198 AATCTTAAGGCTCTTCTCTCTCACTTAAACACTAGCTTCGCTTGACGGAATTTG 257
 168 AATCTTAAGGCTCTTCTCTCTCACTTAAACACTAGCTTCGCTTGACGGAATTTG 227
 258 TTGATGCGGATCAAAATCGGCGGAAATCGGAACTTCACTTCCTGATGAGGCGAC 317
 228 ATCTACCTTCATCCCGGAGAAATCGGAGATTCGCTTCACTTCACTTCACTTCA 281
 318 TCGATACCTTCGCTTCGCTTCGATATCCCTGCTTGACACTTGAATAGATGAC 377
 282 TCGCTTCCTTCACGATATCCCTGCTTGACACTTGAATAGATGAC 341
 378 CTGCTAGATTCAGATTCGATGCGCTTTTATGATGATGATGATGATGATGATGAT 437
 342 CCAGATTCCTTCGCTTCGATATCCCTGCTTGACACTTGAATAGATGAC 401
 438 ATGCAAGACTATTAAGTATCCCGGCTTCGATATCCCTGCTTGACACTTGAATAGAT 497
 402 TCGATATGAAGAACTCTTCAGATTCCTTCGATATCCCTGCTTGACACTTGAATAGAT 458
 498 GCGATACCGGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGAT 557
 459 GCGCTGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGAT 518
 558 TTGATACCTTCGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGAT 614
 519 TTTATACGCTTCGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGAT 578
 615 GCGATTCCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGAT 674
 579 TTTAATCTTCGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGAT 638
 675 ACATTTTGAAGCAATTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGAT 734
 639 ATTTATTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGATGAT 689
 735 CCTGATTCAGCAAGGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGAT 794
 690 TTACCGAGGAGCGGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGAT 749
 795 AAGAAACAAGTACGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGAT 854
 750 AAG-----GCTTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGAT 803
 855 ATGACGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGATGAT 914
 804 GCGATTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGATGATGAT 857
 915 GAATTCAGCAAGGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGAT 974
 858 -----GACCAAGCAAGGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGAT 908
 975 CTTCGAGCGGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGAT 1034
 909 CGATTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGATGATGAT 968
 1035 AAGCAACAAGTACGATATCCCTGCTTGACACTTGAATAGATGATGATGATGATGATGAT 1094
 969 AGAATGCGGCGGAGAGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGATGAT 1028
 1095 GGAGAGCCATTGAAAAGAGGCTTCGATATCCCTGCTTGACACTTGAATAGATGATGATGAT 1154

Db 1029 GCGCGGAGATATAAAAAAGAT---CAACGACAGAGATTTAGAAACGGTGGAGAA 1085
 Oy 1155 TGGTTATCGGAATCTTAATGAATCCCTCAAAAAGATTTCTCGGATTAACCGATCGCT 1214
 Db 1086 TGGTCGCGGAGATTCGTAACCGCTTGCAAGAAATATATTTCTCGGTGACAGATCGAGC 1145
 Oy 1215 AAGTTCGATTCGATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGAT 1274
 Db 1146 AAGTTCGATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGAT 1205
 Oy 1275 TCTGTTGATTAAGCAATG--ATTATGATTCAGTCGAGGATTTGAAAAGCT 1331
 Db 1206 TCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGAT 1265
 Oy 1332 GTGAGATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGAT 1391
 Db 1266 TTGAGGTTGCTTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGAT 1325
 Oy 1392 GAAGCTTTGCTTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGATATTCGAT 1451
 Db 1326 CTGGAATTAATGCTTAATGAATGAT--GTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1384
 Oy 1452 CCAAGTCTTTAGTACACTACCAACCTACTTT 1488
 Db 1385 ATTAAGTGTGATTAAGCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1421

RESULT 8
 AC003027
 LOCUS
 DEFINITION
 Arabidopsis thaliana chromosome I BAC F21M11 genomic sequence.
 AC003027
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 119914)
 Federespiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Luros, S., Schwartz, J., Shin, P., Toriumi, M., Vyotskala, V.S., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Unpublished
 2 (bases 1 to 119914)
 Federespiel, N.A., Palm, C.J., Conway, A.B., Kurtz, D.B., Conway, A.R., Au, M., Araujo, R., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, Y., Ojil, O., Osborne, B., I., Shin, P., Sun, H., Toriumi, M., Vyotskala, V., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (22-OCT-1997) Biochemistry, Stanford University/Stanford Sequencing and Technology Center, 855 California Avenue, Palo Alto, CA 94304, USA
 3 (bases 1 to 119914)
 Federespiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shin, P., Toriumi, M., Vyotskala, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (30-DEC-1998) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
 4 (bases 1 to 119914)
 Federespiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shin, P., Toriumi, M., Vyotskala, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

AUTHORS TITLE JOURNAL	Ecker,J.R. Direct Submission Submitted (26-JAN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE AUTHORS	5 (bases 1 to 101176) Cheuk,R., Shlum,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaverl,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE AUTHORS	6 (bases 1 to 101176) Ecker,J.R. Direct Submission Submitted (15-JUL-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE AUTHORS	7 (bases 1 to 101176) Cheuk,R., Shlum,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaverl,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (19-JUL-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA On Jul 15, 2000 this sequence version replaced gi:2618677. Location/Qualifiers 1..101176 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" /clone="F21B7" join(736..1038,1276..1713,1874..2119,2205..3032, 3117..3274,3483..3610,3725..4134,4313..4435) /note="unknown protein; similar to ESTs dbj AV529515.1, dbj AV529266.1, dbj AV554950.1, dbj AV523761.1, dbj AV523761.1, dbj A1996703.1, and dbj AV546743.1"
COMMENT FEATURES SOURCE	/codon_start=1 /evidence-not_experimental /product="F21B7.1" /protein_id="AAf86508.1" /db_xref="GI:9280639" /translation="MQPLPARSGDHGFNWNHPLLIVAGDETNGTGLGHSQNSCL ARGSSDSKRGDAINPTTVRPSLSHSHYVYLRRSSVCMTIRCSSRVAVGLNQI TCVDATLEAKESVLTVPVQROGTITNVANGYGMAVGPMWLVAASKSMATMKTG LSQDFTSPSSPSSSGSSGSSFMARYMESKOLANGLINGDMGYKTLSTYCQDL PDOSTPASNAIMKVGVGSADENAGAAVADGALVGPFAHSPISALCEPD SGTLTAYSCNNINVOIMPBSRNAPDGSLSESSHVLFLKLHGITSATYODI CFSGOSOWALLISSKGTCHIFVNLSSGSAQAOPCRGEPTLRASSLPMPWTQSLS MOOSSLPPVALSVTRIKTSFGMLNTVSNATPATKVFVPGAAVAAFHKSVTH DIQLNRNALNIHVTPSPGHVHOELLPSVCTEPENGLENVOKTSHVAAVFNKSYH KVPIQMWDYCRSDMLEBERLPKSITEKOYLETVSNHLTSHEDACISLDMSHF EDKYLSCKEPRSHCYLSNFVEVVTGKLPVMONSKISPHWMDSPDSSTGGF EIENVAHAELEIKOKLLPVDPFHSTKATLEDPSMKCYHRSAGSHOVNCIOOI INCNSRPGSTIESSESSEBSTOMENLHSDHMSINSKSLEPYPTVNGITKEIKKN ANCMWRPEVTKLSTLKERTINGETFPPLIDTSVNEQMLSTGKPDKGFGRLHEHC KAAGNPFEBHLKKRIKDEVTNVHHLNVNNNTERKLOGDKWVNQVLNAGVCIITYKDFL RAGGNPFVFFLLIKILRK" complement(4838..6223) /note="hypothetical protein" /codon_start=1 /evidence-not_experimental /product="F21B7.2"
CDS	
AUTHORS TITLE JOURNAL	Ecker,J.R. Direct Submission Submitted (26-JAN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE AUTHORS	5 (bases 1 to 101176) Cheuk,R., Shlum,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaverl,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE AUTHORS	6 (bases 1 to 101176) Ecker,J.R. Direct Submission Submitted (15-JUL-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE AUTHORS	7 (bases 1 to 101176) Cheuk,R., Shlum,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaverl,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (19-JUL-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA On Jul 15, 2000 this sequence version replaced gi:2618677. Location/Qualifiers 1..101176 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" /clone="F21B7" join(736..1038,1276..1713,1874..2119,2205..3032, 3117..3274,3483..3610,3725..4134,4313..4435) /note="unknown protein; similar to ESTs dbj AV529515.1, dbj AV529266.1, dbj AV554950.1, dbj AV523761.1, dbj AV523761.1, dbj A1996703.1, and dbj AV546743.1"
COMMENT FEATURES SOURCE	/codon_start=1 /evidence-not_experimental /product="F21B7.1" /protein_id="AAf86508.1" /db_xref="GI:9280639" /translation="MQPLPARSGDHGFNWNHPLLIVAGDETNGTGLGHSQNSCL ARGSSDSKRGDAINPTTVRPSLSHSHYVYLRRSSVCMTIRCSSRVAVGLNQI TCVDATLEAKESVLTVPVQROGTITNVANGYGMAVGPMWLVAASKSMATMKTG LSQDFTSPSSPSSSGSSGSSFMARYMESKOLANGLINGDMGYKTLSTYCQDL PDOSTPASNAIMKVGVGSADENAGAAVADGALVGPFAHSPISALCEPD SGTLTAYSCNNINVOIMPBSRNAPDGSLSESSHVLFLKLHGITSATYODI CFSGOSOWALLISSKGTCHIFVNLSSGSAQAOPCRGEPTLRASSLPMPWTQSLS MOOSSLPPVALSVTRIKTSFGMLNTVSNATPATKVFVPGAAVAAFHKSVTH DIQLNRNALNIHVTPSPGHVHOELLPSVCTEPENGLENVOKTSHVAAVFNKSYH KVPIQMWDYCRSDMLEBERLPKSITEKOYLETVSNHLTSHEDACISLDMSHF EDKYLSCKEPRSHCYLSNFVEVVTGKLPVMONSKISPHWMDSPDSSTGGF EIENVAHAELEIKOKLLPVDPFHSTKATLEDPSMKCYHRSAGSHOVNCIOOI INCNSRPGSTIESSESSEBSTOMENLHSDHMSINSKSLEPYPTVNGITKEIKKN ANCMWRPEVTKLSTLKERTINGETFPPLIDTSVNEQMLSTGKPDKGFGRLHEHC KAAGNPFEBHLKKRIKDEVTNVHHLNVNNNTERKLOGDKWVNQVLNAGVCIITYKDFL RAGGNPFVFFLLIKILRK" complement(4838..6223) /note="hypothetical protein" /codon_start=1 /evidence-not_experimental /product="F21B7.2"
CDS	
AUTHORS TITLE JOURNAL	Ecker,J.R. Direct Submission Submitted (26-JAN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE AUTHORS	5 (bases 1 to 101176) Cheuk,R., Shlum,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaverl,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE AUTHORS	6 (bases 1 to 101176) Ecker,J.R. Direct Submission Submitted (15-JUL-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE AUTHORS	7 (bases 1 to 101176) Cheuk,R., Shlum,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaverl,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (19-JUL-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA On Jul 15, 2000 this sequence version replaced gi:261

MRG"
 complement(Join(18462..18515,18612..18742,18840..18975,
 19064..19139,19216..19410,19497..19595,19668..19752,
 19865..20103,20179..20253,20345..20649,20798..20902,
 21088..21216,21356..21437,21513..21608,21780..21849,
 21929..22025,22408..22502,22586..22659,22810..22888,
 23041..23101,23206..23315,23411..23556))
 /note="hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="F21B7.7"
 /protein_id="AAF86539.1"
 /db_xref="GI:9280670"
 /translation="MPFKFRSLNFILCPKREKARGOILGYKMKRKAIGFHLSP
 PNIYGHILTAPEVNNHSHOLIRGSGPTAANHNHSSIPESIDVTSVHSFDLIR
 KWTRLNIPGDVPSRACHAAALITGLTILOGIFSGSDVDYVLDMTNKNKIFELV
 GGEFSPRGHMDIAORWLVTEGSGNNGNELLDITMALDTRGPEFMDRLPSGQPS
 GRHVASGSSREDGILFLLCGIDHSGVTGADYVLDITMDSDNWTVPVPAVAPSPRYOHTA
 VFSGSKLHVIGGILNRRLIDGEEAVVADJNTPETASGANGRONQQLMRCHHAASF

Query Match 6.4% Score 109.2; DB 7; Length 101176;
 Best Local Similarity 44.9% Pred No. 5.2e-17;
 Matches 589; Conservative 0; Mismatches 708; Indels 15; Gaps 4;

QY 85 CCCTACCGGTAAACATCTTCGATATCCCTGGTGCACCTTGATTAAGATGACATCCCTTC 144
 Db 37326 CTTCTCTCTCTCACTTCGATGCCCTGGCTCTCTCTCCACCTGCCGATTCCTCT 37385
 QY 145 TGTATTAGACCTTTCGATGCCCAAGACATTTCTTGACACCTGTATTCCTTAATCTTA 204
 Db 37386 TCTTCTCTCTTACCAAAACCTGATGATCTTCTCTCAAGACTTGTGACCAACATCA 37445
 QY 205 AGGCTCTTTTGTCTCTCACTTAAACACTAGTCCGCTTGAGGGAATTTGTTGATGCG 264
 Db 37446 AACATTCCTCTCCATCACTCTCCAACTTCTTCCCTTACGCCGATTAATGATTAATCC 37505
 QY 265 CGATCAAAATCGGGCGCAAAATGCCGAAGTTTCAGTACTCCCTGATGAGGCGCATCGATTA 324
 Db 37506 CGCTCTGCTCC---CGACCCCTCATATTTACACTACACGCTGGCGCAACATCTCTGTTT 37562
 QY 325 CTTTGATGCTTGGCGGAGCTGTGACCAAGATTTTGACTACCTTAAGAGTATCACTGGTAG 384
 Db 37563 TCACCGTATGACAGAGTCTCTGAAACCGACTTGACCACTGAATGATTCCTCCCTAAAG 37622
 QY 385 ATTCATGATTTTGGAGCCCTTTTATGTTATGCCAAGGGTTATAGACATGCAAG 444
 Db 37623 ATATCAGTGTGTGATGAGCGCTTGTGCCAAGTTACCTCTCCACGCTCTCCGGAAG 37682
 QY 445 ACTATTAAGTATCCCGCTCTGACCGGTGACGATACCGCTTTTCTTAACCGTGGCATAG 504
 Db 37683 GAATTCAGATCGAACCAATTAATGGCAAGTCAACCATCTCTCCGAGCTGGAATCT 37742
 QY 505 CCGTGGCTCTGACGCAATCATTCATTTGAGATGCTAAAGTTTGTATATGTTATCA 564
 Db 37743 GTATAGGCAACTGACATACATGTTGATGAGATGAGATCACTTCAATTCATTAAGA 37802
 QY 565 ATGCTTGGCCCTATATTAACAATTTGGGAAAGACGCGGACTTGTTCGCGCAATCTTC 624
 Db 37803 AGTATTTGATGTCAATTAACCAATTCAGCGCAAGATTCGCGCAAGCTTCTTACCTT 37862
 QY 625 TTCCATCTTTTCGATGATGATTCGATTAACAAGTCTGTATGSCCTTAGAGAAACATTTTGA 684
 Db 37863 CTGCGCTATTCACAGCTGCGAAGAAATTAATCAAGAGCCAGAGGAGTACCGCCAGAG 37922
 QY 685 ACGAAATGCAAGATGTTCTTGAATGTTCTCTGATTTGSAAGCAACCCCTCGATTTCA 744
 Db 37923 ATTTGGAGCGGTTTGGAGCAAAACCTGCAAAACATTTCAATGTTACCGCTGAGA 37982
 QY 745 ACAAGGATGACGATACATATGCTCTCTGCTGAAATTCAGAACTTAAGAACAAAG 804
 Db 37983 ACATGGCAAGCCACTTTTACATTTGAGCGGAAGCAGATGAGCAATCTCAAGAGTTGGG 38042
 QY 805 TACTGAATCTCAGAGATCCGAAACGCAATAGCTGTAAAGACGTTCAACAATACGTGTG 864

Db 38043 TTACAGAGC-----AGTGTGAAGATCAATCTCTCTTTCTACCTTCGTGTGACTTAG 38096
 QY 865 GATACCTATGACATGTCATGTCATTAATCAAAAGATACGCTGTATCAGAGAAATCATCGA 924
 Db 38097 CTTTCAATTTGGGTAAACCTTGAATTAAGACACTGTTTCAGACATGTAACAGAGGCCAAGG 38156
 QY 925 ACGACGAAATATGAGCTCGAGTACTTCAGTTTACAGCGGATTCGCGAGACTTCTGACCG 984
 Db 38157 ACGAGTCTTTACCTTGAATGATCAACGTGATTCGAGGAATCTGTTCAAGT---ACACAG 38213
 QY 985 CCCCCTGTCGCCCTAATCTTTCGCAACCTGTCTTGCATGCGCTTCGCAAAAGCAACAC 1044
 Db 38214 AACCAATACCAACAAATCACTTTGGCAACTGTATGCTGCTGATTCATCAGCAAGA 38273
 QY 1045 ATAAAGAGTACTTGGGGATTAAGGCTCTTGTTCGACTTGCACCTTTCAGAGAGCCA 1104
 Db 38274 AACACGATTTGCTAGAGAGAAATATCGTATGCGCTTCAGATGCAATTCACAGCGAGAA 38333
 QY 1105 TTGAAAGAGTTGACCAACGAAAGAGCGTTCTTGACATGCAAAACCTTGTTATCGG 1164
 Db 38334 TCAAGATATGTTATCAAGCG---ATCTGTTGAACAGACACCAAGATGGGACAGAG 38390
 QY 1165 AATCTAATGAAATCCCTTCAAAAGATTTCTGGGATTACCGGATTCGCTTAAGTTGAT 1224
 Db 38391 TACGTAAATGGGTCAATGTCCTCATTAATCCACCTCAATTCAGAGACTCCGAATTTGGGAC 38450
 QY 1225 CGTATGGTGTGATTTTGGATGGGAAAGCCGCAAAATTTGACATTAACCTCTGTTGAT 1284
 Db 38451 TGTATGACATGATTTTGGATTTAGGAAAGCCCTGCAAAATGAATGATGACATTAACAA 38510
 QY 1285 ATGCAAAATTAATTTATGCTGATTCAGTCCAGGATTTTGAAGAAAGTGTGAGATTTGAG 1344
 Db 38511 CAGGTGCTTCTATGCAATTCACAGATCCAGACGCGAGCAATGAGATTGAGATTGGA 38570
 QY 1345 TATCATGCCTAAGATTCATATGATGATGCTATTCGCAAAATCTTTGAAGAAAG 1396
 Db 38571 TAGCACTAGAGAGAAAGAAATGATGTTGTTACTCATCTTTCACAAAGAGG 38622

RESULT 10
 AP000606/c
 LOCUS
 DEFINITION
 Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MTO24,
 complete sequence.
 ACCESSION
 AP000606
 VERSION
 AP000606.1 GI:6045161
 KEYWORDS
 HTG.
 SOURCE
 Arabidopsis thaliana (strain: Columbia) DNA, clone: lib: M1tsu1 P1
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euPhylohytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 1 (sites)
 REFERENCE
 1 Nakamura, Y.
 2 Structural Analysis of Arabidopsis thaliana Chromosome 3. II
 JOURNAL
 Unpublished (1998)
 2 (bases 1 to 82360)
 AUTHORS
 Nakamura, Y.
 TITLE
 Direct Submision
 JOURNAL
 Submitted (13-OCT-1999) to the DDBJ/EMBL/GenBank databases.
 Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
 Gene Structure 2; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 (E-mail: ynakamura@kazusa.or.jp, URL: http://www.kazusa.or.jp/gene-s2/
 Tel: 81-438-52-3935, Fax: 81-438-52-3934)
 FEATURES
 source
 1. 82360
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"

FEATURES	*	159496	166884	contig of 7389 bp in length.
Location/Qualifiers				
source		1..166884	/organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="7"	
BASE COUNT		55357 a	28890 c 29784 g	52486 t 367 others
ORIGIN				

Query Match	4.28;	Score 71.8;	DB 60;	Length 166884;
Best Local Similarity	46.28;	Pred. No. 1.7e-07;		
Matches 604;	Conservative	0;	Mismatches 627;	Indels 76; Gaps 8;

OY	187	CTGTTATCCCTAACTTCTTAAGCGCTCTTGTGTCCTCACTCTAATAAGACTACGTCCGGCTTA	246
Db	112029	CTCTCATCTCTAAACTTAAAGTCTCTCTCTCTCAGACCACGCTCTCTCCACGTATCTCCACTCG	111976
OY	247	GCGGAAMATTGTGGATGCCGATCAATCGGGGGAATG-CGSAAGTTTAGTACTCCGCTG	306
Db	111969	CCGGCCCCCTCGTGIGGGATTCATTAACAACMAA---CATTCAATGTCTATTTCTCTG	111913
OY	307	ATGAGGGCGACTGATPAACTTGTATGTTGGGAGCTGTACCGAGATTTTGACTACTTA	366
Db	111912	ACGACAAAGATGCCGCTGATCTCACGGTAGCGATCTACAGGTACGGTAGACCTTTCCATTCTT	111853
OY	367	AAGCTCATCAACTGATAGATATCCAAATGATTTGCATGGCGCTTTTTTATGTTATGCGACGGG	426
Db	111852	CCGGAGATGAACCTCGTCCCGCAGAGATTTCATTCTTTTAGTCCCTGAGTTAAC-----	111796
OY	427	TTAATAAGSACATGCAAGAAGCTATAAGATGCCGCTGTAGCCGTCGAAAGTAAACGTTT	486
Db	111797	-----GGTTTCCGACGAATACGCTCGGGTGTCTGCGCGTTACAGTAAACGTTT	111751
OY	487	TTCCTTAACCGTGGCATAGCCGCGGCTCTGAGCGCATCATTTCAATTGGCAGATGCTAAA	546
Db	111750	TCCCAAACAAGATTTCTCTCGCGCTAACCGCATCATAGCCGTTTATGATGTGAAAA	111691
OY	547	GTTTGTATGTTCATCATGCTTGGGCGCTATATAACAATTTGGGAAAGACGCGACT	606
Db	111690	CACACGCACTGTTTCTCAAGATGggc-----TCACAACCTATAACAAGAACAAAGAAG	111637
OY	607	TGTTGTCCGCGAATCTTCTCCATCTTTTCATGTAGATGATATACAAAGATCTGATAGGC	666
Db	111636	CTCTGCGACACGATCTAGTACCGTCTTATGATGTATCATTTGTCAAAGCCACAGGGAC	111577
OY	667	TAGAGAAACATTTTGSAGAAATGCAAGATGTTCTTGAATGTTCTAGATTGGAA	726
Db	111576	TCGAGCAAAACTCTTGACCGGTGATATCACCTTCCAACAACCAACCACTTAAAC	111517
OY	727	-----GCAAAACCCCTCGATTCAACAGGTACGAG-TTACATATGTGCTCTCCCTTG	777
Db	111516	TGTTTTCTTCCAAATAATATGATCCGATTTCTCGAGTCAAGTACCGGTTAACTCGAG	111457
OY	778	CTGAATATCCAGACTTAAGAACAAGTACTGAATCTCAAGAGATCCGAACCGACATATAC	837
Db	111456	AAGACATTAANAACCTA---CGTGAACGATGGAGACCAAGTACACACGCTAAAGCAGCTCC	111400
OY	838	GTGTAACGACTTCAACATGACGTGTGGATATGACATGACATGGTGCAAATCAAMAAG	897
Db	111399	GTTTGTGCACTTTGTCTATTACTTAAAGCTTACGTTATATCAATGATGGTGAAGATGCGTG	111340
OY	898	ATGACGTGCTATCAGAGGATCATTCGAACGAGAAATGAGCTCGAGTACTTCAGTTTAA	957
Db	111339	GAG-----GTGACCGGACCCGGTTGTGTGCTGCTTGGTTTG	111304
OY	958	CAGCGGATTCCGAGAGACTTGTACGCCGCCGCTGTCCGCTTAACACTTGTGGCAAGTGC	1017
Db	111303	CATCAGACTTTCGGTACAGGTTAAACCTTCGCTTACGCCCGCAGTGTCTTGGGAACTCGA	111244
OY	1018	TTG-----CGTCATGCGTTGCAAAAGCAACACATCTAAAGATTTAGTTGGGGATTA	1066

Db 111243 TTGTGGGGTGGGAGATTTTCGACGTGMAAGCAACCAATATTGGAGAGAGAGAA 11118
Qy 1067 AGGCGCTTCTTGTTCGAGTTGCAGCTATTGGAGAGCAATTGAAAGAGCTTGCACMACGA 11126
Db 111183 AAGGTTTCATCACGGGGGAGAGACTTTTACCGGATGGGTCAACGAGATTGTGTCGGAGA 1111224
Qy 1127 AAAAGCGCTTCTTGCAGATGCAGAAAACCTTGGTTATCGGAATCTAATGGAATCCCTTCAMA 1186
Db 111123 ATATAGAGAAGAATATATGCTTTTACGTTTGAAGCGTTTAAGAGATGGAACCGGMAAGC 111066
Qy 1187 AACATTTCGGGATTTACCGGATGCGCTAAAGTGCATTCTATGCTGTACATTTTGGATG 1246
Db 111063 AATGATATACG - TTGCGGGTGCACCCGGTTAGGAGATTTACGGGCTGCTGATTTTGGGCTG 111060
Qy 1247 GGGAAAGCCTGCAGAAATTTGACATTACCTCTGTGATTATGCAGAAATGATATTATGAT 1306
Db 111005 GGGAAACCGGTCAGAGGTAGAGATTGTGACGATTGATTAAGATGCTTCGGTTTCATTGTC 110946
Qy 1307 TCAGTCCAGGAGATTTTGAAGAAAGGTGGAGATATGAGATATCATTTGCTTAAGATTCATAT 1366
Db 110945 GGAAGATGGAGATGSAATGCTGCTGTGAGAGTTGGTGTGTTGAAGMAACATATGT 110886
Qy 1367 GGATGATTTGCAGAAAATCTTTGAAGAAGCGCTTTGCTCTTTGTCATAGTCTCTTAAATA 1426
Db 110885 GGAACGTTTGTGTCATTATTTCTCATCGGGTGGATTAATAATGAGAACGTTTGGCTTC 110826
Qy 1427 GAACCATTTTGTGCAATTAAGATCAAGACGCTTTTAGTAACTAC 1473
Db 110825 TAGTTTATGTGTGTTTATTTAATTCACACATCATCTTTAATATTCATTC 110779

Query Match	3.88;	Score 64.2;	DB 7;	Length 66237;
Best Local Similarity	49.48;	Pred. No. 1.3e-05;		
Matches 222;	Conservative 0;	Mismatches 223;	Indels 4;	Gaps 2;

QY	948	TTCAATTTTACAGCGGATTTGCCGAGACTTTCGACGCCCGGTGCGGCTTAACACTTT	1007
Db	26239	TACGATTCGCGTGTGGATTTGTTCGAAGCCTTATGTTCCACCGGTTCCATCCAGTATTTT	26358
QY	1008	GGCAACTGTCTTGGCTCATGCCGTTGCCAAAAGCAACACATTAAGACTTATGGGGATPAA	1067
Db	26359	GGTATTTGTGTATCGCTTGG--TTTAAATGTGCTTAAAGCGGAGACCTTTTATGAGTAA	26417
QY	1068	GGCGTCTTGTTCGAGTGGACGCTTTGGAGAAGCATTTGAAAAGAGTTGCACACGA	1127
Db	26418	GAAGGCTTTTAAAGCTGCGTCGAGAAATGTTATGTAATCTGTGGAGCAATG--GATAG	26474
QY	1128	AAAGCGTCTTCTTCGAGATGCACAAAACCTTGTTATGGAATCTAATGGAATCCCTTCAAA	1187
Db	26475	AATGTTCGATTTGAAGATTCACAGAGATTTTGGAGAAGGTTTACGACCTTTCACACAGAA	26534
QY	1188	AGATTTCTCGGGGATTAACCGATGCCCTAATGTTCGATTGATGTGTAGATTTTGGATGG	1247
Db	26535	CAGGTTTTGTCTGTGGTCCGCGGTGCACCGGTTCCGAGTTTACGGGTGAGATTTTGGGTGG	26594
QY	1248	GGAAAGCCTGCACAAATTTGACATTTACCTCTGTGTTATGCGAATTTATATATGTGATT	1307
Db	26555	GCTAGACCGCGAAGAAAGTGTGTGTGTGTCATATGACCAAGGGAAGCCATTTCTTTTCCG	26654
QY	1308	CAGTCCAGCGGATTTTGAAGAAAGGTGTGGAGATTGGAGTATCATTTGCTTCAGATTCATAT	1367
Db	26655	GAGACTAGAGATGGGAGTGTGGGTGTAGAGCTTGGCTTCTCCCTCAGAAACATGAATNG	26714
QY	1368	GATGCATTTGCCAAAATCTTTGAGAAGG	1396
Db	26715	GATGTTCTGTGTGATTTGCTTCACAAGGG	26743

RESULT	13
LOCUS	AC069470/c
DEFINITION	Arabidopsis thaliana chromosome I clone IGF-F7F7, *** SEQUENCING INFORMATION
ACCESSION	AC069470
VERSION	AC069470.8
KEYWORDS	GI:9558603
SOURCE	HTG; HTGS; PHASEL.
ORGANISM	thale cress. Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE	1 (bases 1 to 136047) Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Ronning,C.M., Koo,H., Fujii,C.Y., Uterback,T., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M. Arabidopsis thaliana 'IGF' BAC 'F7F7' genomic sequence near marker 'm172'
TITLE	Unpublished
JOURNAL	2 (bases 1 to 136047)
REFERENCE	Town,C.D. and Kaul,S.
AUTHORS	Direct Submission
TITLE	Submitted (01-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@igf.org
JOURNAL	On JUL 28, 2000 this sequence version replaced gi:9440543.
COMMENT	* NOTE: This is a 'working draft' sequence. It currently * consists of 4 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 100163: contig of 100163 bp in length * 100164 100213: gap of unknown length * 100214 112623: contig of 12410 bp in length * 112624 112673: gap of unknown length * 112674 116366: contig of 3963 bp in length

FEATURES	Source	Location/Qualifiers
* 116637	116686: gap of unknown length	
* 116687	136047: contig of 19361 bp in length.	
BASE COUNT	46028 a 22872 c 22213 g 44783 t 151 others	
ORIGIN		
Query Match	3.7%; Score 63; DB 59; Length 136047;	
Best Local Similarity	45.1%; Pred. No. 2.9e-05;	
Matches 436; Conservative	0; Mismatches 520; Indels 11; Gaps 5;	
1. 136047	/organism="Arabidopsis thaliana"	
	/db_xref="taxon:3702"	
	/chromosome="1"	
	/clone="IGF-F7F7"	
QY 460	CGCTGTGACCGCTGCAAGTACACCGTTTTCCTAACCCGTGGCATFACCGGTGGCTTGACGG	519
Db 28857	CGGTCTCTTCTTCTTGGAAATCACTTTGTTTCCCCACCAAGGGTTTGTATGTGGTATCGCTT	28798
QY 520	CACATCATTTCAATTCAGATGCTATAAGTTTGTATGTTCATCATGCTTTGGGCTATA	579
Db 28797	CTCATCATCTCTGTATATGACGGGAAACAGTGGTTATAGTTTATAAATCATGGCTCACA	28738
QY 560	TTAACAAATTTGGGAAGACGGGACTTGTGTCCGGCAATCTTCCATCTTTCATAT	639
Db 28737	TTTGTAAACATGGA-----GCCATGGAATTTACCGAGATTTGACTCCGGTTTATGATC	28684
QY 640	GATCGATATCAAAAGCATGTGATGGCCATAGAGAAACATTTTGGAAACAAATGCAAGAT	699
Db 28683	GTACGGTTATCATATGTCTCCCTGCAAGTCTTGTATGCAAAAATCATTTACTTTTGTCTATTT	28624
QY 700	TTCTGTAATGTCTCTAGATTTGGGAAGCAAAACCCCTCGATTTCAACAAGTACAGCTTA	759
Db 28623	TCTCAGAGTGAAGATGATGTTTCATGATCATTTGAAGCTCTTCTCCCAAGAGATCAGCC	28564
QY 760	CATATGTCTCTCCCTTCTGTAATCCAGAAAGCTAAAGAACTACTGATCTTCAGAG	819
Db 28563	CCGACTTGTAGTCAGATCTGCTCGAGTTTGACTCGAGAGAAATTTAG-AGAATCTTAGGAG	28505
QY 820	GATCCGAACCGCAATACGTGTGAACGAGCTTCACATACGCTGATACATGATGACAT	879
Db 28504	CAAGCCAAAGAGGAGTACAGCTGCTCTCATCATATGAGCTTCACTGTCCAAAGTTCGTTGTT	28445
QY 880	GCATGCTCAAAATCAAAAGATGACGTCTGTATCAGAGGAATCATCGAACGCAAAATGAC	939
Db 28444	GCCAAACGGGTATCTTTGGACATGCTTGGTAAAGACGGCTGGAGA--GATGGAACAGAC	28387
QY 940	TGCACTACTTCAAGTTTACAGCGAGTTGCGCGAGACTTCTGACGGCCCCGCTTCGGCTTA	999
Db 28386	CGGTCCGTTTCAATGATATGCTGCTGTATTTTCAGAACCGGTTAGATCCACCGGTTCTTGACA	28327
QY 1000	ACTACTTTTGGCACTGTCT-TGCGTCATGCGTTTGGCAAAAGCAACATTAAGAGTTAGTT	1058
Db 28326	TGTATTTTGGGAACGCTGTTCCCGATAGGTGCTTTGGTTACAAAGCGAATGTTT	28267
QY 1059	GGGGTAAAGGGCTCTGTGTGCACTGTGCACTATTTGGAAGAACCATTTGAAAGAGTTTG	1118
Db 28266	TGGGGAAGATGGGTTTGTATAATATGATGCAAAATTTCTAAGTGATTTGCTTACAAAGTATG	28207
QY 1119	CACACAGCAAAAGCGGTTCTTTCAGATGCAAAACCTTGATGTTATCGAATCTTAATGGAATC	1178
Db 28206	GTTTACGAAGTTGGAACAATTTGCGAATGTTCATTAATGCAACAAAGATGTCGAAC	28147
QY 1179	CCTTCAAAAGATTTCTGCGGATTACCGGATCGCTTAAGTTCGATTCGTATGCTGTAGAT	1238
Db 28146	CAGGTACACAGC-TTGGATTCATTTCCCGGTTCCGAACAGTTTGGGCTATACGGGTCAGAT	28088
QY 1239	TTTGGATGGGGAACCCGCAAAATTTTACATTTACCTCGTTGGATTATGACGAATTTGATTT	1298
Db 28087	TTTCGGGTGGGGAACCAATGTAACAGTGAATGCTGTCAATTTGACCGAAAGAGCGCTTT	28028
QY 1299	TATGTGATTCAGTCCAGGATTTTGAAGAAAGCGTGTGAGATTTGAGATATCATTTGCCCTAAG	1358

	a	c	g	t	others
BASE COUNT	46028	22872	22213	44783	151
ORIGIN					

Query Match	3.7%;	Score 63;	DB 59;	Length 136047;
Best Local Similarity	45.1%;	Pred. NO. 2.9e-05;		
Matches 436;	Conservative 0;	Mismatches 520;	Indels 11;	Gaps 5;

OY	460	CGCTGTAGCCGTGACAGTAACCGTTTTTCCATACCGGGCATTCACCGGCTCGACGG	519
Db	28857	CGGTTCTTTCTTTCGCAAGTCACTTTGGTCCCAACCAAGGGTTTGATGGTATCGCTT	28798
OY	520	CACATCATTTCAATTCAGATGCTAAAGTTTTTGTAAATGTTCAATCAATGCTTGGGCTATA	579
Db	28797	CTCATCATTTCTGTATTTGAGACGGGAAACAGTGGTTAGGTTTATAAATCATGGGCTACA	28738
OY	580	TTAACAAATTTTGGGAAACCGGGACTGTGTCCGCCAATCTTCTTCATCTTTCGATA	639
Db	28737	TTTGTAAACATGGA-----GCCATGATTTTAAACGAGAAATTTGACTCCGGTTTTAGATC	28684
OY	640	GATCGATATCATAAAGATGCTGTATGGCTTAGAGAAACATTTTGGAAACGAATATGCAAGATG	699
Db	28683	GTACGGTTATCAATGATTCCTCGTAAGTCTTGATGCAAAATATCATGTACATTTTGTCTATATT	28624
OY	700	TTTCTGAAATGTCTCTAGATTTTGGACCAAAACCCCTCGATTTCAACAAGTACGACTTA	759
Db	28623	TCTCGAAGTGAAGATGATTTTCAATATCATATGTAAGACTCTTCCCTCCCAAGGAGATCAGCC	28564
OY	760	CATATGTCCCTGCTCCCTTGTCTGAAATCCAGAAAGCTTAAGAACAAAGTACTGAATCTCAGAG	819
Db	28563	CCGACCTAGTACGAGATCTCGCTCCGAGTTGACTCGAGAGAAATPAG-AGAACTTTAGGGAG	28505
OY	820	GATCCGAACCGACAATACGTGTGTAACGACGTTTCACAATGACGTGTGATACGTATGAGACAT	879
Db	28504	CAAGCCAAAGAGGAGTCAAGCTGCTCTCATCTATGAGCTTCACTTGTCAACGTTCCTGTT	28445
OY	880	GCATGCTCAAAATCAAAAGATGACGTCGTATCTAGAGAAATCATCGAAGCAACGAATATGACG	939
Db	28444	GCCAAAGCGGTATCTTTGACATGCTGTGTGTAAGAACCGCGGAGGA--GATGGAACAGACG	28387
OY	940	TGCAATCTTCAATAGTTTACAGCGGATTTCCCGAGACTTCTGACGCCGCCGTTGCCGCTTA	999
Db	28386	CGGTCCGTTTCAATGATGCTGCTGATTTTCAGGAACCGGTAGATCCACCGTTCCTGAGA	28327
OY	1000	ACTACTTTGGCACTGTCT-TGCGTCAATGCTTGSCAAAAGCAACATTAAGATTACTT	1058
Db	28326	TGTATTTTGGGGAACGCTGTTCCCGAATAGGTTGCTTTGGTACAAAGCGAATGTGTTTT	28267
OY	1059	GGGGATTAAGGCTCTGTGTGCATTTGCAGCTATTTGGAGAAGCCATTGAAGAAGGCTT	1118
Db	28266	TGGGGAAGATGGGTTTCTTAATATGGTGCGAATTTCTAAGTGATTCGTTGGAAGTATTG	28207
OY	1119	CACAACGAAGAAAGCGTCTTTCGACAGATGCAAAACCTTGATTCGGAATCTTAATGGAATC	1178
Db	28206	GTTTACGAAGTTGGAACAATTTTCGAAATGTCAATTAAATGCAACAAAGATGTGAAC	28147
OY	1179	CCTTCAAAAAGATTTCTGCGGATTACCGGATGCGCTAAGTTGCATTCGTATGSGTATAGT	1238
Db	28146	CAGGTACACAGA-TTGGATTCATTTCCCGGTTCCGAACCAAGTTGGGCTATACGGGTCACAT	28088
OY	1239	TTTGGATGGGGAAGCCATCGCAAAATTTTGCATTTACCTCGTTGATTATATGCAATTTGATT	1298
Db	28087	TTTCGGGTGGGGAACCAATGTACAGTATAGATTTGGCTCATTTGACCGAAAGCAGGCGTTT	28028
OY	1299	TATGTGATTCAGTCCAGGGAATTTTGAAGAAAGGTGTGAGATTTGAGATATCATTTGCCCTAAG	1358

Db 28027 TCGATGTCGAGAGAGGATGAACCCGAGCTGTGAGATCGCTTGTGTTGAAGAA 27968
 Qy 1359 ATCTATATGATGATTTGCAAAAATCTTTAAGAGGCTTTGCTCTTGTATAGTCT 1418
 Db 27967 TGTGAGATGGACATATTTATCTTTCATAAATGCTTGAAGATGCTTCTTAGGCT 27908
 Qy 1419 CTTTAAAT 1425
 Db 27907 CTTTGAT 27901

RESULT 14
 AB028618
 LOCUS
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, p1 clone:MOD1,
 complete sequence.
 ACCESSION AB028618 GI:5041971
 VERSION AB028618.1
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsids.
 1 (sites)
 REFERENCES
 AUTHORS Nakamura,Y.
 TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. II
 JOURNAL Unpublished (1999)
 REFERENCE 2 (bases 1 to 85690)
 AUTHORS Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUN-1999) to the DDBJ/EMBL/GenBank databases.
 Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
 Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
 (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
 Fax:+81-438-52-3934)
 FEATURES
 source location/Qualifiers
 1..85690
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="MOD1"
 /clone_lib="Mitsui P1"
 BASE COUNT 27956 a 15199 c 15751 g 26784 t
 ORIGIN

Query Match 3.6%; Score 61.8; DB 7; Length 85690;
 Best Local Similarity 56.7%; Pred. No. 5.6e-05;
 Matches 114; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1199 GATTACCGGATCGCTAAGTTCGATTCGATGCTGATGATTTGGATGGGAAAGCCTGC 1258
 Db 16287 GATAGCCGGGTGACCCGGTTGGAGATATACGAGCGGATTTGGGGGAGAACCCCG 16346
 Qy 1259 AAAATTGACATACCTCTGTTGATATGCAAGATGATTTATGTCAGTCCAGGGA 1318
 Db 16347 TAAAGTTGATATTTGTCATTCATGACCAAGAGAGGATCGCAATGCTGAGAGAGCTGA 16406
 Qy 1319 TTTTGAAGAAAGTGTGAGATGATGATATGATGCTTAAGATTCATATGATGATTTGC 1378
 Db 16407 TGAGTCAGGTGGGCTGATGATGATGATGTTGAAAAAGACGAAATGATTCGTCGT 16466
 Qy 1379 AAAATCTTTGAAGAGCCTT 1399
 Db 16467 GTCCTTTTTCACAAATGCTTT 16487

RESULT 15

AC069470
 LOCUS
 DEFINITION Arabidopsis thaliana chromosome I clone IGF-F7F7, *** SEQUENCING IN
 PROGRESS ***, 4 unordered pieces.
 AC069470
 ACCESSION AC069470.8 GI:95858603
 VERSION AC069470.8
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 136047)
 Lin,X., Kaul,S., Town,C.D., Beilto,M., Creasy,T.H., Ronning,C.M.,
 Koo,H., Fujii,C.Y., Utterback,T., Barnstead,M.E., Bowman,C.L.,
 White,O., Nierman,W.C. and Fraser,C.M.
 Arabidopsis thaliana 'IGF' BAC 'F7F7' genomic sequence near marker
 'm172'
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 136047)
 REFERENCE Town,C.D. and Kaul,S.
 AUTHORS Direct Submission
 TITLE Submitted (01-JUN-2000) The Institute for Genomic Research, 9712
 JOURNAL Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org
 COMMENT On Jul 28, 2000 this sequence version replaced 91:9440543.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 100163: contig of 100163 bp in length
 * 100164 100213: gap of unknown length
 * 100214 112623: contig of 12410 bp in length
 * 112624 112673: gap of unknown length
 * 112674 116536: contig of 3963 bp in length
 * 116537 116686: gap of unknown length
 * 116687 136047: contig of 19361 bp in length.
 FEATURES
 source location/Qualifiers
 1..136047
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="IGF-F7F7"
 BASE COUNT 46028 a 22872 c 22213 g 44783 t 151 others
 ORIGIN

Query Match 3.6%; Score 61.8; DB 59; Length 136047;
 Best Local Similarity 56.7%; Pred. No. 5.8e-05;
 Matches 114; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1199 GATTACCGGATCGCTAAGTTCGATTCGATGCTGATGATTTGGATGGGAAAGCCTGC 1258
 Db 91188 GATAGCCGGGTGACCCGGTTGGAGATATACGAGCGGATTTGGGGGAGAACCCCGT 91247
 Qy 1259 AAAATTGACATACCTCTGTTGATATGCAAGATGATTTATGTCAGTCCAGGGA 1318
 Db 91248 TAAAGTTGATATTTGTCATTCATGACCAAGAGAGGATCGCAATGCTGAGAGAGCTGA 91307
 Qy 1319 TTTTGAAGAAAGTGTGAGATGATGATATGATGCTTAAGATTCATATGATGATTTGC 1378
 Db 91308 TGAGTCAGGTGGGCTGATGATGATGATGTTGAAAAAGACGAAATGATTCGTCGT 91367
 Qy 1379 AAAATCTTTGAAGAGCCTT 1399
 Db 91368 GTCCTTTTTCACAAATGCTTT 91388

Search completed: October 20, 2000, 03:24:05

Job time: 6483 sec

• • • • •

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 01:35:37 ; Search time 1792.79 Seconds
(without alignments)
5873.147 Million cell updates/sec

Title: US-08-894-356C-1
Perfect score: 1703
Sequence: 1 TCATTATGAGCAATCCAA.....AAAAAAAAAAAAAAAAAAAA 1703

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estpl1:*
71: em_estpl2:*
72: em_estpl3:*
73: em_estpl4:*
74: em_estpl5:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

117: gb_gss18:*
118: gb_gss19:*
119: em_gss13:*
120: gb_gss20:*
121: gb_gss21:*
122: gb_gss22:*
123: gb_gss23:*
124: gb_gss24:*
125: em_gss14:*
126: em_gss15:*
127: em_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122.8	7.2	591	BE434257	BE434257 EST405335
2	109.8	6.4	606	AW221049	AW221049 EST297518
c 3	107.6	6.3	603	AW616119	AW616119 EST296884
4	107.2	6.3	542	BE436185	BE436185 EST407263
5	101.6	6.0	497	AW930934	AW930934 EST356777
6	101	5.9	583	AW221050	AW221050 EST297519
7	97.4	5.7	549	AI490445	AI490445 EST248771
8	93.6	5.5	687	AW102336	AW102336 s086006.y
9	91.6	5.4	404	AW737239	AW737239 EST338666
c 10	91.6	5.4	543	AW650650	AW650650 EST329104
11	90.8	5.3	518	AW650280	AW650280 EST328734
12	85.4	5.0	526	AL386584	AL386584 MLCB35E10
c 13	85.2	5.0	705	AW2399055	AW2399055 EST305729
14	84.6	5.0	788	BE187619	BE187619 EST336180
15	82.8	4.9	487	AW616206	AW616206 EST307245
16	75.8	4.5	451	BE433920	BE433920 EST404998
c 17	74.6	4.4	605	AW440958	AW440958 AV440958
18	74.2	4.4	550	AL380276	AL380276 MLCB51E01
19	73.8	4.3	488	BE461227	BE461227 EST412646
20	70.2	4.1	485	AL386583	AL386583 MLCB35E10
21	69.2	4.1	569	AW719572	AW719572 LJNEST6H1
22	68	4.0	561	AW831571	AW831571 sm34h07.y
c 23	61.8	3.6	346	Z37266	Z37266 ATTS4009.pe
24	60.2	3.5	668	BE323055	BE323055 NF001A07P
25	60	3.5	490	AW038515	AW038515 EST280198
26	59.4	3.5	409	AW093479	AW093479 EST286659
c 27	58.6	3.4	615	AW5867016	AW5867016 EST318639
28	51.2	3.0	656	AW692527	AW692527 NF056D05S
29	49	2.9	597	AW774948	AW774948 EST334099
30	48	2.8	311	BE353747	BE353747 EST355090
31	47.8	2.8	367	AW203420	AW203420 sf30a09.y
c 32	47.8	2.8	434	AV340186	AV340186 AV540186
c 33	47.4	2.8	313	AW523212	AW523212 AV523212
c 34	47.4	2.8	359	AI755098	AI755098 cr35d10.x
c 35	47	2.8	1101	CNS001VP	AL078809 Drosophil
36	46.8	2.7	433	AU038087	AU038087 AU038087
37	46.8	2.7	658	AW695291	AW695291 NF093E09S
38	46.8	2.7	677	AW586972	AW586972 EST318595
c 39	46	2.7	1289	B09480	B09480 T24H11-T7 T
c 40	45.6	2.7	463	AA430523	AA430523 zw21b07.s
c 41	45.6	2.7	625	AW310044	AW310044 sf30a09.x
c 42	45.4	2.7	350	CNS00WUX	AL094071 Arabidops
43	45.2	2.7	607	C92395	C92395 C92395 Dict
44	45	2.6	1101	CNS002M0	AL062875 Drosophil
c 45	44.8	2.6	461	BE440122	BE440122 HTM1-926F

ALIGNMENTS

RESULT 1
BE434257

LOCUS BE434257 591 bp mRNA EST 24-JUL-2000
DEFINITION EST405335 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone cLEG15H6, mRNA sequence.
ACCESSION BE434257
VERSION BE434257.1 GI:9432100
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 591)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL Unpublished (2000)
COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@clemson.edu
5 prime sequence.
FEATURES
source
1..591
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG15H6"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKM; Uadapt: Site_1: EcoRI; Site_2: XhoI; Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of the fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."
BASE COUNT 159 a 129 c 121 g 182 t
ORIGIN

Query Match 7.2%; Score 122.8; DB 35; Length 591;
Best Local Similarity 53.7%; Pred. No. 4.3e-22;
Matches 303; Conservative 0; Mismatches 252; Indels 9; Gaps 2;
QY 132 ATGCAGTCCCTTCGTGTTTACGACTTTCCGTTACCCAGAACACACATTTCTTGACACACTGTT 191
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 ATGAGCGGATATATATTTTACAGCTCTCCATTTTCAACCCGATTTTCGTTCAAAACAT 66
QY 192 ATCCCTAATCTTAAGGCTCTTTTCTCTCACTCTAAACACTACGTTCCGCTTAGCGGA 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 ATTCCTCTCTTAAATTCATCTCTCCCTCACTCTCAACACTATACGCCCTTAGCCGA 126
QY 252 AATTGTTGATGCGCAATAATCGGGGAATCGCGAAGTTTTCAGTACTCTCCGCTGATGAG 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 AACGTTGCTTCCACATAGATACAAAGGATATCTGAGTTACGTTA-----TGTGACA 180
QY 312 GGCACATCGATACTTGTGCGGAGTCTGCACAGGATTTTGACTACCTTAAAGGT 371
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GGAGATCTGTGCTGTACTTTTTCGAGACTCATATGAATTTTCAATTAATCTCATTTGGT 240
QY 372 CATCACTGGTAGATTCGAATGATTCGATGGCTTTTTTATTTATTTATGTTATGCCACGGTTATA 431
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 GACCATCGCGTAGGCTAAGGATTTTAT---CACTTTGTCTCCTAGTAGGGAACCT 297
QY 432 AGGACCATGAAGACTATATAGTGATCCGCTCTAGCCGTGCAAGTAACCGTTTTTCCT 491
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 AAGGATCGACGGGGGTCACACTAGCCCGCTCTTAGCCATTCAGGTGACACTTTTTCCG 357

```

QY 492 AACCGTGGCATAGCGTGGCTGTGACGGCAGCATCATTCATTCAGATGCTAAAGTTT 551
  || || || || || || || || || || || || || || || || || || || || ||
Db 358 ATCTTGTTGATCATCTGGTTTCTACTAACCATCATCTGTTGTTGATGGAGCTACTATA 417
  || || || || || || || || || || || || || || || || || || || || ||
QY 552 GTAATGTTTCATCAATGCTTTGGCCCTATATTAAACAAATTTGGGAAAGACGGCGACTTGGT 611
  || || || || || || || || || || || || || || || || || || || || ||
Db 418 GCAGGGTTCATTAAGGCGTGGCTCTACTCCACAAATTCGGTGACATGAACAATTCCTTA 477
  || || || || || || || || || || || || || || || || || || || || ||
QY 612 TCCGGGAATCTCTTCATCTTTTCGATAGATCGATATCAAGATCTGTATGGCCTAGAG 671
  || || || || || || || || || || || || || || || || || || || || ||
Db 478 TCGAATGAGCTAAATTCATTTATGATAGTCCGTAGTAAAGACCCATATGGCAAGG 537
  || || || || || || || || || || || || || || || || || || || || ||
QY 672 GAAACATTTTGGCAAGCAATGCAA 695
  || || || || || || || || || || || || || || || || || || || || ||
Db 538 ATGTCATCTGGGAAGAATGAAA 561
  || || || || || || || || || || || || || || || || || || || || ||

RESULT 2
AW221049 606 bp mRNA EST 07-DEC-1999
LOCUS EST297518 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone cLEF3F1, mRNA sequence.
ACCESSION AW221049
VERSION AW221049.1 GI:65327733
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanales; Solanales; Solanales; Solanales; Solanales;
1 (bases 1 to 606)
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F.,
Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M.,
Fraser, C.M., Martin, G.B., Tanksley, S.D., and Giovannoni, J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
Location/Qualifiers
1. .606
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF3F1"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site1: EcoRI; Site2:
XhoI, cLEF. Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT 147 a 143 c 123 g 193 t
ORIGIN

Query Match 6.4%; Score 109.8; DB 20; Length 606;
Best Local Similarity 52.2%; Pred. No. 1.2e-18;
Matches 321; Conservative 0; Mismatches 282; Indels 12; Gaps 3;

QY 19 AAATCGTGAAGTCTTGGAAATCCCAAGTTACACCACCTCTGACACAAACAGATGTCG 78
  || || || || || || || || || || || || || || || || || || || || ||
Db 2 ACATGGCCACCGTATGATGGAATGCAATGTCAGTTGCCACCTCCCGCGCGCCACGAGG 61
  || || || || || || || || || || || || || || || || || || || || ||
QY 79 AGTTATCTACCGGTACATCTTCGATATCCCTCGTTCACCTTGTAATAGATGCAGT 138
  || || || || || || || || || || || || || || || || || || || || ||

```

```

Db 62 TGATA---CTCCCTCTTACTTATTTTACCATGTTTGGTAGGGTTTCGCCGTATGAGC 118
  || || || || || || || || || || || || || || || || || || || || ||
QY 139 CCCTTCCTGTTTACGACTTTCCTGACCAAGAACACACATTTCTTTGGACACTGTTTCCCTA 198
  || || || || || || || || || || || || || || || || || || || || ||
Db 119 GGATATTTATTTTACAAGCTCTCCATTTTCAACCCGATTTTCGTTCAAAACATTTATTCCTC 178
  || || || || || || || || || || || || || || || || || || || || ||
QY 199 ATCTTAAGGCTCTTGTCTCTCACCTCTAAACACTACGTTCCGCTTAGCGGAAATTTGT 258
  || || || || || || || || || || || || || || || || || || || || ||
Db 179 CTCCTTAAATTCACCTCTCCCTCACCTCTCAACACTATACGCCCTTAGCGGAAACGTTG 238
  || || || || || || || || || || || || || || || || || || || || ||
QY 259 TGATGCCGATCAATCGGCGAAATGCCGAAGTTTTCAGTACTCCCGTATGAGGCGCACT 318
  || || || || || || || || || || || || || || || || || || || || ||
Db 239 CTGTGTCACCTAGATACAAACGGATATCTCGATGTTACGTTA-----TGTGACAGGAGATT 292
  || || || || || || || || || || || || || || || || || || || || ||
QY 319 CGATAACTTTGATCGTTGCGGAGTCTGACGAGATTTTGACTACCTTAAAGGTCATCAAC 378
  || || || || || || || || || || || || || || || || || || || || ||
Db 293 CTGTGCTGTACTTTTTCGAGACTGATGAATTTCAATTTATCTCATTTGTTGACCATC 352
  || || || || || || || || || || || || || || || || || || || || ||
QY 379 TGTGATATTCCAAATGATTTGCGCTGCTTTTATCTGTTATGCCACGGTTATAAGGACCA 438
  || || || || || || || || || || || || || || || || || || || || ||
Db 353 CGGTAAGGCTAAGGATTTTAT---CACTTTGTTCTTAAGTTAGGGGAACCTAAGGATG 409
  || || || || || || || || || || || || || || || || || || || || ||
QY 439 TGCAAGACTATAAAGTATCCCGCTCGTAGCCGTGCAAGTAAACCGTTTTCCTAACCGTG 498
  || || || || || || || || || || || || || || || || || || || || ||
Db 410 CACCGGGGGTCCAACTAGCCCGCTCTTAGCCATTTCAGCTGACACTTTTCCGAATCTTG 469
  || || || || || || || || || || || || || || || || || || || || ||
QY 499 GCATACCGTGGCTCTGACGGCACATCAATTCATTCAGATGCTAAAGTTTGTAAATGT 558
  || || || || || || || || || || || || || || || || || || || || ||
Db 470 GTGATCCATTTGTTTCACTAACCATCATGTTGTTGGTATGGAGCTACTATAGAGGTT 529
  || || || || || || || || || || || || || || || || || || || || ||
QY 559 TCATCAATCTTTGGGCTATATTAAACAAATTTGGGAAAGACGGGACTGTTTGTCCGGA 618
  || || || || || || || || || || || || || || || || || || || || ||
Db 530 TCATTAAAGCGTGGGCTCTACTCCACAAATTCGGTGGACATGAACANTCTTATCGAATG 589
  || || || || || || || || || || || || || || || || || || || || ||
QY 619 ATCTTCTTCCATCTT 633
  || || || || || || || || || || || || || || || || || || || || ||
Db 590 AGCTAATTCATTTT 604
  || || || || || || || || || || || || || || || || || || || || ||

RESULT 3
AW616119 603 bp mRNA EST 24-MAR-2000
LOCUS EST296884 L. hirsutum trichome, Cornell University Lycopersicon
DEFINITION hirsutum cDNA clone cLHT6C7 3', mRNA sequence.
ACCESSION AW616119
VERSION AW616119.1 GI:7322152
KEYWORDS EST.
SOURCE Lycopersicon hirsutum.
ORGANISM Lycopersicon hirsutum.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanales; Solanales; Solanales; Solanales; Solanales;
1 (bases 1 to 603)
van der Hoeven, R.S., Bezzarides, J.L., Matern, A.L., Holt, I.E., Liang
F., Hansen, T., Craven, M.B., Bowman, C.L., Ronning, C.M., Niernan, W.,
Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.
Generation of ESTs from wild tomato (Lycopersicon hirsutum)
trichomes
Unpublished (2000)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
3 prime sequence.
Location/Qualifiers
1. .603
/organism="Lycopersicon hirsutum"
/db_xref="taxon:62890"
/clone="cLHT6C7"

```



```

Qy 379 TGTAGATTCCAAATGATTGTCATGCCCTTTTATGTTATGATCCACGGGTATTAAGGACCA 438
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 353 CGCGTAAGCGTAAGGATTTTAT---CACTTTGTTCCTAAGTAGGGGAACCTAAGGATG 409
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 439 TGCAGACTATAAGTGATCGCGTCGTAGCGGTGCGCAAGTAACCGTTTCTTACCGGTG 498
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 410 CACCGGGGTCCAACTAGCCCGCTCTTAGCCATTACAGTGCACACTTTTTCGGAATCTTG 469
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 499 GCATAGCGGTGCTGTGACGCGCATCATCAATTGCAGATGCTAAAGTTTCTTAATGT 558
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 470 GTGATCACTGGTTTCACTAACCATCATGTTGTTGGTGAGGAAGTACTATAGCATGTG 529
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 559 TCATCAATGCTTGGGCCCTATATTACAAAT 589
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 530 TCATTAAAGCGGTGGCTCTACTCCACAAAT 560
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
AI490445 549 bp mRNA EST 29-JUN-1999
LOCUS EST248771 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
DEFINITION cLED21J4, mRNA sequence.
ACCESSION AI490445
VERSION AI490445.1 GI:4385755
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 549)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
,C.L., Nieman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU.
FEATURES
Location/Qualifiers
1..549
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED21J4"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 176 a 91 c 126 g 155 t 1 others
ORIGIN

Query Match 5.7%; Score 97.4; DB 11; Length 549;
Best Local Similarity 53.2%; Pred. No. 2.4e-15;
Matches 284; Conservative 0; Mismatches 232; Indels 18; Gaps 3;

Qy 522 CATCATTCATATGCAGATGCTAAAGTTTGTGAATGTTTCATCAATGCTTGGCGCTATATT 581
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9 CATCATGTTGTTGGTGAGGACTACTATACGAGGTTTCATTAAAGCGGTGGGCTCTACTC 68
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 582 AACAAATTTGGAAGACCGGAGCTGTTGTCGCGGAATCTCTTCCATCTTTTCATAGA 641
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 69 CACAAATTCGGTGACATGAACAAATCTTATCTGAATGAGCTAATTCCTATTTTATGATAGG 128
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 642 TCGATAATCAAGATCTGTATGGCTAGAGAAACAATTTTGGAAAGAAATCAAGATGTT 701
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 TCCGTAGTAAAGAACCCCATATGGACAAAGGATGTCATCTGGGAAGAAATCAAAAGCAAG 188
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 702 CTTGAATGTTCTCTAGATTGGGAAGCAACCCCTCGATTCAACAGGTACGAGCTACA 761
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 189 AATCTAGATATGCGTGACGTT---ATGACTCTCTCTGAACACAGGTTTCGAGGTACA 245
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 762 TATGTCCTCTCCCTTGTGTAATCCAGAAGCTAAAGAACAAAGTACTGAATCTCAGAGGA 821
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 246 TTTACTATTAAAGAGAGATGAATAGAGAAACTGAAGAATTTTATATTGAACCTCAAGAAGA 305
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 822 TCCG-----AACCGACAATAGCTGAACGAGCTTCACAANTGACGTGTGGATAC 869
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 306 CGGGTAGTAGTAGTACTACTCTAACTCATCTCTTTTACTGTAAACGAGTGTAT 365
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 870 GTATGGGACATGCATGGTCAAAATCAAAAGATGACGCTGTATCAGAGGAATCATCGAACGAC 929
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 366 ATATGGACTGTTTGTATAAATC---AGAGACCGGATAGGAGAAGAGATCATATAGAT 422
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 930 GAAATAGCTCGAGTACTTTCAGTTTACAGCGGATGTCGAGGACTTCTGACGCCGCCG 989
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 423 GATAGTGAATGGAATGTTTCGGATGTGCAGCAGATTTTAGAGCGCGATTCAATCCACCA 482
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 990 TGTCGGCTAACTACTTGGCAACTGCTTCGCTCATCGTTGCAAAAGCAACA 1043
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 483 CTTCCTCAATCTTATTTTGGGAATGTCATAGTTGGGTATGTCACANAATCAATA 536
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
AI102336 687 bp mRNA EST 06-DEC-1999
LOCUS sd86406.y1 Gm-cl009 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl009-1284 5' similar to TR:Q9ZWB4 Q9ZWB4 F21M11.13 PROTEIN. ;
mRNA sequence.
ACCESSION AI102336
VERSION AI102336.1 GI:6072949
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 687)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corvelli,V., Khanna
,A., Bollu,B., Marra,M., Hillier,L., Rucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,X., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 431.
FEATURES
Location/Qualifiers
1..687
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl009-1284"

```


/clone_lib="Cm-cl009"
 /lab_host="XL10-Gold"
 /note="vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; The mRNA was isolated from entire roots of 2-month-old 'Williams' plants that were greenhouse grown in 5-gallon pots. To suppress nodulation, Black gold All-purpose potting soil was supplemented with: 0.36g/L available phosphoric acid (P205), 20mg/L urea N, 0.16g/L S, 0.49mg/L B, 2.5mg/L Cu, 0.15g/L Fe, 13.53mg/L Mn, 0.26mg/L Mo, 14mg/L Zn, 20mg/L Ca, and the following nutrients in a slow-release form (Osmocote): 0.165g/L ammonia N, 0.185g/L nitrate N, 0.35g/L available phosphoric acid, and 0.35g/L soluble potash. No nodules were visible on the roots at harvest. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stratagene's first-strand synthesis primer was used [GAGAGAGAGAGAGAGACTGCTCGAG(T)-18]. After second-strand synthesis is, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 400bp cutoff, using a SizeSeph 400 Spin column from Pharmacia. The column eluent was then ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. This library was constructed by Dr. Paul Kelm and Dr. Virginia Corvelli."

Query Match	5.5%	Score 93.6;	DB 20;	Length 687;
Best Local Similarity	49.1%	Pred. No. 2.6e-14;		
Matches 302;	Conservative 0;	Mismatches 304;	Indels 9;	Gaps 2;
Qy 24	GTGAAGGTTCTTGAANAATGCCAAGTTTACACCACCATCTGCACACAACAGATGTCGAGTTA	83		
Db 58	GTGAAGATCTTTAGAGCAAGTGAAGTTGGTCCACCACCACATGC---TCACCTTCCCTCAACC	114		
Qy 84	TCGCTACCGGTAAACATTCTTCGCATATCCCTGGTTGGCACTTGAATAAGATGCAGTCCCTT	143		
Db 115	ACCTTCCCTCACTTCTTCGATATTCCATGGTTCTACTGCCACCATCCACGGATC	174		
Qy 144	CTGTTTTACGACTTTCGGTACCCCAAGAACACATTCTTGGACACTGTTATCCCTAAATCTT	203		
Db 175	TTCTTCTATGACTTTTCCGCCACACACACACTTTCTCCAACACGCACTTCCCATCTCTC	234		
Qy 204	AAGGCTCTTTGTCTCTCACACTCTAAAACACTACGTTTCGGCTTACGGGAAATTTGTTGATG	263		
Db 235	AAACACTCTCTTTCCCTAAACCCTCAACACTCTTCTGCCCTTCTCCTCCAATCTCATCGTT	294		
Qy 264	CCGATCAAAATCGGCGGAAATGCCGAAGTTTCAGTACTCCCGTGATGAGGGCGACTCGATA	323		
Db 295	CCTCCACAGAACCCCATCTCTTCACATACGCTAC-----CTTGATGGAGACTCTCTC	348		
Qy 324	ACTTTGATCGTTGGGGAGTCTGACACGAGATTTTGACTACTCTTAAAGGTCACTCAACTGGTA	383		
Db 349	TCCTTCAACGTTGCAGAGTCCACCGACACTTCACCCCTCCTCCATCATGATTCACCACAA	408		
Qy 384	GATTCCAATGATTTGCATGGCCCTTTTTTATGTTATGCCACGGGTTATAAGGACCATTGCAA	443		
Db 409	GACGTTCCAAATTTGGCACCCCTCTTGTTCGCCGCTTTCTTACCCCAACGCTGTGACCAAGAT	468		
Qy 444	GACTATAAAGTGATCCCGCTCGTAGCCGTGCAAGTAACCGTTTTTCTTAACCGTGGCATA	503		
Db 469	GGCCAGCTGTGTTCCCTCTTATGGCCATTCAGGTTACAAATTTTCCCAAAGCTGTGGCTTC	528		

QY	504	GCCTGGGCTCTGACGGCACATCATCAATTCAGATGCTAAAGTTTTTGTAAATGTTTCATC	563
Db	529	ACCATATGCTCTCACCTTCAACCACTTGCAGCAGCGCAAGTCACTTCCACCATTTTCATC	588
QY	564	AATGCTTGGGCTATATTAACAAATTTGGGAAAGACGGGACTTGTGTGTCGGCGAATCTT	623
Db	589	AAGTCTGGGCTCTCTTGTGCANAGCANNAGAAACATGGCTTNNCTTTCAACCTCTNTG	648
QY	624	CTTCATCTTTTCGAT	638
Db	649	TCTCTTCTTCTCAT	663
RESULT	9		
AW737239			
LOCUS	AW737239	404 bp	EST 25-APR-2000
DEFINITION	EST338666	tomato flower buds, anthesis, Cornell University	
ACCESSION	AW737239	Lycopersicon esculentum cDNA clone cTOD2D1 5', mRNA sequence.	
VERSION			
KEYWORDS			
SOURCE	AW737239.1	GI:7646184	
ORGANISM	Lycopersicon esculentum		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
	1. 404		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA496"		
	/db_xref="taxon:4081"		
	/clone="cTOD2D1"		
	/clone_lib="tomato flower buds, anthesis, Cornell University"		
	/tissue_type="flower"		
	/dev_stage="anthesis"		
	/note="vector: pBlueScript SK(-); Site.1: EcoRI; Site.2: XhoI; supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."		
BASE COUNT	125 a	52 c	96 g 131 t
ORIGIN			
Query Match	5.4%	Score 91.6;	DB 24; Length 404;
Best Local Similarity	56.8%	Pred. No. 7.8e-14;	
Matches 216;	Conservative 0;	Mismatches 149;	Indels 15; Gaps 2;
QY	1040	AACACATAAAGAGTTAGTTGGGGATAAAGGCTTCTTGTTCAGTTCAGCTATTGGAGA	1099
Db	12	AAGCATTTTACATAGTTAGTGAAGGAGGCTTTAAATTCGGGTAGAAATCAATTGGAGA	71
QY	1100	AGCCATTCAAAGAGTTGCAACAGAAAAGGCGTTCTTCAGATGCAAAAACCTTGTT	1159
Db	72	AGTCATTCAAGAAAATGAAGGATGATGAATGGTCTCTTAATGGTGA-----TTGGTT	125
QY	1160	ATCGGAATCTAATGGAATCCCTTCAAAAAGAAATTTCTCGGGATTACCGGCTCGCCTAAATGTT	1219

[illegible]

```

Db 126 AAAAGTATACGACAAATGATGATTCGATCATATTTTCATATGCTGGATCGCAAAAACA 185
Qy 1220 CGATTTCGTATGCTGTAGATTTTGGATGGGAAAGCCTGCAAAATTTGACATTAACCTCTGT 1279
Db 186 TGACTTATATGCTGCTGATTTTGGATGGGAAAGCCGCAAGTTGGAAATTCATTTCCAT 245
Qy 1280 TGATTATACGACA-----ATTGATTTATGTGATTCAGTCCAGGATTTTGAAGG 1330
Db 246 TGACAATGATGATGCGGCTATTCGATCTCTTAGTAAATCTAAAGATTTTGATGGAGA 305
Qy 1331 TGTGGAGATTGGAGTATCATTCCTTAAGATTCATATGATGCATTTGCAAAAATTCITGA 1390
Db 306 TTTAGAGATTGGTTGTGTTTGTCTAAACCTCGAATGAATGCTTTGCTTCTATATTCAC 365
Qy 1391 AGAAGGCTTTTGCTCTTTGT 1410
Db 366 TCACGAGCTTAGCTTTCTAT 385

RESULT 10
AW650650/c 543 bp mRNA EST 04-APR-2000
LOCUS EST329104 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION esculentum cDNA clone cLEI13N5 5', mRNA sequence.
ACCESSION AW650650
VERSION AW650650.1 GI:7411888
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 543)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source Location/Qualifiers
1. 543
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEI13N5"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/notes="vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
XhoI; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
BASE COUNT 186 a 114 c 75 g 168 t
ORIGIN

Query Match 5.4%; Score 91.6; DB 23; Length 543;
Best Local Similarity 56.8%; Pred. No. 8.4e-14;
Matches 216; Conservative 0; Mismatches 149; Indels 15; Gaps 2;

Qy 1040 AACACATAAAGATAGTTGGGATAAAGGCTTCTGTCGACGTTCAGCTATTGGAGA 1099
Db 526 AAGGCATGTTGCTAGTTGGAGGAAGGCTTTAAATTCGCGTAGAATCAATTTGGAGA 467

```

```

Qy 1100 AGCCATTGAAAAGAGGTTGCACAAAGCGGCTTTTCAGATGCAAAAACCTTGTT 1159
Db 466 AGTCATTCAAGAAAAAATGAAGGATGATGAATGGTCTCTTAATGGTGA-----TTGGTT 413
Qy 1160 ATCGGATCTAATCGAATCCCTTCAAAAGATTTCTCGGGATTTACCGGATCGCCTAAGTT 1219
Db 412 AAAAGTATACGACAAATAGATGTGATTCGATCATTTTCAATTCGTGATCGCAAAAACA 353
Qy 1220 CGATTTCGTATGCTGTAGATTTTGGATGGGAAAGCCTGCAAAATTTGACATTAACCTCTGT 1279
Db 352 TGACTTATGCTGCTGATTTTGGATGGGAAAGCCGCAAGTTGGAATTCATTTCCAT 293
Qy 1280 TGATTATACGACA-----ATTGATTTATGTGATTCAGTCCAGGATTTTGAAGG 1330
Db 292 TGACAATGATGATGCGGCTATTCGATCTCTTAGTAAATCTAAAGATTTTGATGGAGA 233
Qy 1331 TGTGGAGATTGGAGTATCATTCCTTAAGATTCATATGATGCATTTGCAAAAATTCITGA 1390
Db 232 TTTAGAGATTGGTTGTGTTTGTCTAAACCTCGAATGAATGCTTTTGTCTATATTCAC 173
Qy 1391 AGAAGGCTTTTGCTCTTTGT 1410
Db 172 TCACGAGCTTAGCTTTCTAT 153

RESULT 11
AW650280 518 bp mRNA EST 04-APR-2000
LOCUS EST328734 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION esculentum cDNA clone cLEI12F13 5', mRNA sequence.
ACCESSION AW650280
VERSION AW650280.1 GI:7411518
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 518)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source Location/Qualifiers
1. 518
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEI12F13"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/notes="vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
XhoI; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
BASE COUNT 134 a 119 c 100 g 165 t
ORIGIN

Query Match 5.3%; Score 90.8; DB 23; Length 518;
Best Local Similarity 51.1%; Pred. No. 1.4e-13;

```


Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE
AUTHORS
1 (bases 1 to 705)
VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.

TITLE
JOURNAL
COMMENT
ESTs from roots of Medicago truncatula after Rhizobium inoculation
Unpublished (1999)
Contact: VandenBosch K
Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel: 409 845 7707
Fax: 409 845 2891
Email: kate@mail.bio.tamu.edu

TIGR sequence name:MTAZ86TK
More information is available at: . (and for clone ordering info)
http://chryslie.tamu.edu/medicago
Seq primer: Skmod (CTA gAA CTA gTg gAT CC).

FEATURES
Location/Qualifiers

source
1..705
/organism="Medicago truncatula"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="KV2-1104"
/clone_lib="KV2"
/tissue_type="Seedling roots"
/dev_stage="2 days post-inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain SOLR"
/note="vector: pBluescript SK -; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in SOLR cells."

BASE COUNT 240 a 172 c 76 g 217 t
ORIGIN
Query Match 5.0%; Score 85.2; DB 21; Length 705;
Best Local Similarity 52.6%; Pred. No. 4.5e-12;
Matches 239; Conservative 0; Mismatches 203; Indels 12; Gaps 2;

QY 946 ACTTCAGTTTACAGCGGATTCGCGAGGACTTCTGACGCCCGCTGTCGGCTTACTACT 1005
|||||
DB 637 ACTTTGGATTTCATGCTGGTGTATACCAAGATTGGAGTATCAAGTACCAAGAATTATT 578
QY 1006 TTGGCAACTGTCTTGCATCGTTCGTTGCAAAAGCAACACATAAAGAGTTAGTTGGGGATA 1065
|||||
DB 577 TTGGTAATGTGTGGATTGGAAGGGTATCATTCACAAAAGAGCTTACTAGAGAAG 518
QY 1066 AAGGCTTCTTTGTTCAGTTTCAGCTATTGGAGAGCCATTGAAAGAGGTTGCACACG 1125
|||||
DB 517 ATGTTATTGTTTGCACCTAAAGAAATTGGAAGCACAATTAAAGAACTT-----G 467
QY 1126 AAAAAGGCGTCTTCAGATCGAAAACTTGGTTATCGGAATCTAATGGAATCCCT---T 1182
|||||
DB 466 ATGCATCAATTTTGGAGAGGGTGAAGAATGGATTTTGGATTGGGAAATGTTATAGGGT 407
QY 1183 CAAAAGAAATTTCTCGGGATTACCGGATCGCCTAAGTTCGATTCGTATGGTCTAGATTTTG 1242
DB 406 CAGAGGAACATGTTTCATGTTACATGGTCTCCAAATTAAGAACTTATGAATGAATTTG 347
QY 1243 GATGGGAAAGCCGTCAAAATTTGACATTACCTCTGTTGATTATGCAGAAATTGATTATG 1302
DB 346 GGTGGGGAGCCCTAAGAAATAGAGGAGTTTCAATTGATTTTACRAGAGGTGTTCTT 287
QY 1303 TGATTCACTCAGGGATTTTGAAGAGGTGGAGATTGGAGATTCATTCATTCCTAAGATTC 1362
|||||

Db 286 TTGTTGAGTAGAGATTTTGAGGTGGATTGAGATAGGTAGCTTTGCTTAAGAGTA 227

QY 1363 ATATGGATGCATTTGCCAAAATCTTTGAAGAGG 1396
|||||

Db 226 AATGGATATTTTCACCTTTTCTTCAAGAATGG 193
|||||

RESULT 14

LOCUS

BE187619 788 bp mRNA EST 22-JUN-2000
DEFINITION EST336180 KVO Medicago truncatula cDNA clone pkV0-1613, mRNA sequence.

ACCESSION

BE187619

VERSION

BE187619.1 GI:8666803

KEYWORDS

EST.

SOURCE

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE

AUTHORS

1 (bases 1 to 788)
VandenBosch,K., Hur,J., Beremand,P., Peng,H. and Ellis,L.

TITLE

ESTs from uninoculated seedling roots of Medicago truncatula

JOURNAL

Unpublished (1999)

COMMENT

Contact: VandenBosch K

Department of Biology

Texas A&M University

College Station, TX 77843-3258, USA

Tel: 409 845 7707

Fax: 409 845 2891

Email: kate@mail.bio.tamu.edu

TIGR sequence name:MTGAU50TK

More information is available at: .

http://chryslie.tamu.edu/medicago

Seq primer: Skmod (CTA gAA CTA gTg gAT CC).

FEATURES

Location/Qualifiers

source

1..788

/organism="Medicago truncatula"

/cultivar="genotype Al7"

/db_xref="taxon:3880"

/clone="pkV0-1613"

/clone_lib="KV0"

/tissue_type="Seedling roots"

/dev_stage="Immediately prior to inoculation with Sinorhizobium meliloti (0 hour)"

/lab_host="E. coli strain SOLR"

/note="vector: pBluescript SK -; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in SOLR cells."

BASE COUNT 237 a 188 c 100 g 262 t

ORIGIN

Query Match 5.0%; Score 84.6; DB 34; Length 788;

Best Local Similarity 51.5%; Pred. No. 6.7e-12;

Matches 254; Conservative 0; Mismatches 224; Indels 15; Gaps 2;

QY 84 TCGTACCGGTAACATTTCTCGATATCCCTGGTTCGACTTGAATAGATCAGTCCCTT 143
|||||

Db 5 TCCCTCCCTTAACATTTCTTGACATTTCTTGGCTTCTTTCTCTCCAAAGCCACCCTA 64
|||||

QY 144 CTGTTTTACGACTTTCCTACCCCAAGACACATTTCTTGGACACTGTTATCCCTAATCTT 203
|||||

Db 65 TTTTTCATGAATTTCTCTCATTTCCATTTCCCATTTTCCAAACACCATTTGCCAAACTT 124
|||||

QY 204 AAGCCTCTTGTCTCTCAGTCTTAAACACATCAGTTCGGCTTACGGGAAATTTGTTGATG 263
|||||

Db 125 AAACAATCTCTCTCTCTCACACTTCAACATTTACTTCCCATTTCCGGGTACATT----- 177
|||||

Matches	248;	Conservative	0;	Mismatches	227;	Indels	9;	Gaps	2;
Qy	57	CCATCTGACACACAGATGCGAGCTTATCGCTACCGGTAAACATCTTTCGATATCCCTGG	116						
Db	2	CCACCTCCGACGCGCAACGGAGTGGTACTCTCTACTTATTTTGACCATGTTTGG	61						
Qy	117	TTGCACCTTGAAATAGATGCGAGTCCCTTCTGTTTACGACTTTCGGTACCCCAAGAACACAT	176						
Db	62	TTAGGGTTTCACCTTATTTAGCGGATATATTCTACAAGCTCTCCATTTCCAAATCCGAT	121						
Qy	177	TTCTTTGACACTGTTATCCCTTAATCTTAAAGCCCTCTTTTGCTCTCAGCTCTAAAACACATAC	236						
Db	122	TTCTGTTCAAAACATATATTCTCTCTTAAAAAATTCACCTTTCCCTTCACCTCTCAACACATAT	181						
Qy	237	GTTCCGCTTAGCGGAAATTTGTTGATCCGATCAAAATCGGGGAAATCGCGAAGCTTTTCAG	296						
Db	182	ATGCCCTTAGCCGGAACGTTGCTGTTCCTACTAGATACAAAGGATATCTTGAGTTACGT	241						
Qy	297	TACTCCCGTGATAGGGCGACTCGATAACTTTGATCTGTGGGAGCTGACCCAGAGATTTT	356						
Db	242	TA-----TGTGCAGGAGATTCTGTGCTGTTAAATTTTCTGAGACTGATATGGATTTTC	295						
Qy	357	GACTACCTTAAAGTGCATCACTCGTAGATCCAAATGATTTGCATGGCCCTTTTATATGTT	416						
Db	296	AATTTATCTCATTTGGCGACCATCCGCGTAATGCTAAGGATTTTAT---CACTTTGTTCTCT	352						
Qy	417	ATGCCACGGGTTATAAGGACCATCGAAGACTATAAAGTGATCCCGCTCGTAGCCGTCCAA	476						
Db	353	AAGTTTAGGGGAACCTAAGGATGACCCCGGGGTCCAAATAGCCCCGCTCTTAGCCATTCAA	412						
Qy	477	GTAACCGTTTTTTCCTAAACCGTGGCATAGCCGCTGGCTCTGACGGCACATCATTCAAATGCA	536						
Db	413	GTGACACTTTTTCGATCTTGGTGTATCCATTGGTTTCACTAACCATCATGTTGTTGGT	472						
Qy	537	GATG 540							
Db	473	GATG 476							

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 46)	van der Hoeven, R. S., Bezzerides, J. L., Matern, A. L., Holt, I. E., Liang, F., Hansen, T., Craven, M. B., Bowman, C. L., Ronning, C. M., Nierman, W., Fraser, C. M., Martin, G. B., Gvannoni, J. J., and Tanksley, S. D.	Generation of ESTs from wild tomato (<i>Lycopersicon hirsutum</i>)

```

3 prime sequence.
FEATURES
  source
    Location/Qualifiers
      1. .487
        /organism="Lycopodium hirsutum"
        /db_xref="taxon:62890"
        /clone_lib="L. hirsutum trichome, Cornell University"
        /clone_id="LHRTD15"
        /tissue_type="trichome"
        /dev_stage="mixed stages"
        /note="Leaves of various stages were shaken in liquid
nitrogen, shearing off trichomes. This procedure yields
a mixture of cells which is highly enriched for trichome
cells, likely with minor contaminations of other types of leaf
cells."

```

Query Match 4.9%; Score 82.8; DB 23; Length 487;
Best Local Similarity 51.2%; Pred. No. 1.8e-11;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 03:24:05 : Search time 3710.09 Seconds
(without alignments)
1791.724 Million cell updates/sec

Title: US-08-894-356c-6
Perfect score: 1522
Sequence: 1 MTGACCACTCTCTCGAATC.....TTTTTTTTTTTTTTTTTTTT 1522

Scoring table: IDENTITY_NMC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenDbml:.*
1: gb_bal:.*
2: gb_bal:.*
3: gb_bal:.*
4: gb_bal:.*
5: gb_bal:.*
6: gb_bal:.*
7: gb_bal:.*
8: gb_bal:.*
9: gb_bal:.*
10: gb_bal:.*
11: gb_bal:.*
12: gb_bal:.*
13: gb_bal:.*
14: gb_bal:.*
15: gb_bal:.*
16: gb_bal:.*
17: gb_bal:.*
18: gb_bal:.*
19: gb_bal:.*
20: gb_bal:.*
21: gb_bal:.*
22: gb_bal:.*
23: gb_bal:.*
24: gb_bal:.*
25: gb_bal:.*
26: gb_bal:.*
27: gb_bal:.*
28: gb_bal:.*
29: gb_bal:.*
30: gb_bal:.*
31: gb_bal:.*
32: gb_bal:.*
33: gb_bal:.*
34: gb_bal:.*
35: gb_bal:.*
36: gb_bal:.*
37: gb_bal:.*
38: gb_bal:.*
39: gb_bal:.*
40: gb_bal:.*
41: gb_bal:.*
42: gb_bal:.*
43: gb_bal:.*

44: em_htg1:.*
45: em_htg2:.*
46: em_htg3:.*
47: em_hum5:.*
48: gb_pl3:.*
49: gb_pr5:.*
50: gb_htg8:.*
51: gb_htg9:.*
52: gb_htg10:.*
53: gb_htg11:.*
54: gb_htg12:.*
55: gb_htg13:.*
56: gb_htg14:.*
57: gb_in3:.*
58: gb_htg15:.*
59: gb_htg16:.*
60: gb_htg17:.*
61: em_htg4:.*
62: em_htg5:.*
63: em_htg6:.*
64: em_htg7:.*
65: em_hum6:.*
66: gb_htg18:.*
67: gb_htg19:.*
68: gb_htg20:.*
69: gb_htg21:.*
70: gb_htg22:.*
71: gb_htg23:.*
72: gb_v11:.*
73: gb_v12:.*
74: gb_bal:.*
75: em_htg8:.*
76: em_htg9:.*
77: em_htg10:.*
78: em_htg11:.*
79: em_htg12:.*
80: em_htg13:.*
81: em_htg14:.*
82: em_htg15:.*
83: em_htg16:.*
84: em_htg17:.*
85: em_htg18:.*
86: em_htg19:.*
87: em_htg20:.*
88: em_htg21:.*
89: em_htg22:.*
90: em_htg23:.*
91: gb_pr6:.*
92: gb_pr7:.*
93: gb_sts1:.*
94: gb_sts2:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	765.4	50.3	1479	5	E12756
2	763.6	50.2	1476	7	AB029340
3	193	12.7	1679	7	AB010708
4	193	12.7	1703	5	E12753
5	172.6	11.3	1508	5	E12757
6	153.8	10.1	1622	5	E12754
7	153.8	10.1	1622	5	AB026494
8	106.4	7.0	66237	7	AB016892
9	102.2	6.7	101176	7	AC002560
10	87.8	5.8	119914	7	AC003027
11	77.8	5.1	85690	7	AB028618
12	77.8	5.1	136047	59	AC069470

C	13	69.6	4.6	82360	8	AP000606
C	14	69.6	4.6	166884	61	AC074226
C	15	65.8	4.3	224448	70	PFMAL426
C	16	63	4.1	11829	32	AE001425
C	17	61.6	4.0	136047	59	AC069470
C	18	60.6	4.0	5481	57	PFSEPTIG
C	19	60.6	4.0	201508	54	AC026290
C	20	59.6	3.9	110000	51	AC023048-0
C	21	59.6	3.9	175463	71	CNS05PEO
C	22	59.2	3.9	4362	57	DD060170
C	23	59.2	3.9	4617	57	SFAJ5398
C	24	59.2	3.9	85690	7	AB028618
C	25	59.2	3.9	110000	52	AC023789-1
C	26	59.2	3.9	234914	55	AC027647
C	27	59	3.9	6372	57	DDICNPA
C	28	59	3.9	77835	71	PFMAL13P2-3
C	29	58.8	3.9	83110	71	PFMAL13PB
C	30	58.8	3.9	144334	30	AC004153
C	31	58.8	3.9	173003	52	AC024245
C	32	58.4	3.8	1598	57	DMYOSLC
C	33	58.4	3.8	2426	57	PFU53326
C	34	58.4	3.8	146882	39	AC010933
C	35	58.2	3.8	110000	52	AC023789-1
C	36	58	3.8	4458	57	DDIDP87
C	37	58	3.8	241536	70	AC005140
C	38	57.6	3.8	46013	70	AP001873
C	39	57.6	3.8	174640	42	AC017044
C	40	57.4	3.8	14529	32	AE001397
C	41	57.2	3.8	975	57	DD087514
C	42	57.2	3.8	197294	39	AC010989
C	43	57	3.7	7218	5	I66494
C	44	57	3.7	157984	30	AC004688
C	45	57	3.7	196149	30	AC004709

ALIGNMENTS

RESULT	1					
LOCUS	E12756	1479 bp	DNA	PAT	24-JUN-1998	
DEFINITION	Perilla ocimoides mRNA for acyltransferase, partial cds.					
ACCESSION	E12756					
VERSION	E12756.1	GI:3251588				
KEYWORDS	JP 1997070290-A/4.					
SOURCE	unidentified.					
ORGANISM	unclassified.					
REFERENCE	1 (bases 1 to 1479)					
AUTHORS	Ashikari, T., Tanaka, Y., Fujiwara, H., Nakao, M., Fukui, Y., Yonekura, K., Mizutani, M. and Kusumi, T.					
TITLE	GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY					
JOURNAL	Patent: JP 1997070290-A 4 18-MAR-1997.					

COMMENT	OS	Perilla ocimoides
	PN	JP 1997070290-A/4
	PD	18-MAR-1997
	PF	30-JAN-1996 JP 1996046534
	PR	17-FEB-1995 JP 95P 67159, 29-JUN-1995 JP 95P 196915 PI
		ASHIKARI TOSHIHIKO, TANAKA YOSHIKAZU, FUJIWARA HIROYUKI, PI NAKAO MASAHIRO,
	PI	FUKUI YUKO, YONEKURA KEIKO, MIZUTANI MASAKO, KUSUMI TAKAKI PC
		C12N15/09, A01H1/00, C07H21/04, C07K14/42, C12N9/10, (C12N9/10, PC
		C12R1:865),
	PC	(C12N9/10, C12R1:19):
	CC	strandedness: Double;
	CC	topology: Linear;
	FH	Key
	FT	source
	FT	1. 1479
	FT	/organism='Perilla ocimoides' FT
	FT	/tissue_type='leaves
	FT	/clone='psAT208'

FT	CDS	1. 1343
FT		/product='acyltransferase'.
FEATURES		
source	1. 1479	location/Qualifiers
		/organism='unidentified'
		/db_xref='taxon:32644'
BASE COUNT	420 a 316 c 331 g 411 t	1 others
ORIGIN		
Query Match	50.3%; Score 765.4; DB 5; Length 1479;	
Best Local Similarity	71.1%; Pred. No. 4.8e-146;	
Matches 1056; Conservative 1; Mismatches 417; Indels 11; Gaps 3;		
8	CCCTCTCGAATCCCGGAGTGGCGCCGCTCCAGCAGCGGAGGCGTGCAGTACGCC	67
1	CCGTGATCGAAGAGTGTAGAGTGGCGCCGCGGACGTGGGCGGCAATCGCTGC	60
68	CGCTACCTCTTCGACATGAGGTGGCTCATTTCCACCCCATGCTTCAGCTTCTCT	127
61	CGCTACATTCCTCGACATGAGGTGGCTCATTTTCATCCATGCTTCAGCTCTCTCT	120
128	ACGACTCCCTCTTCGAAACCGCTTCTCGAAGCGTCTTCGGAACCTCAACAT	187
121	ACGAATCCCTCTTCGAAACCAATTTTTCAGAAATCATCTCTTCGAAACCAAT	180
188	CCCTATCTCAACCTCAACCTCTTCCTTCGATCATCATCATCATCATCATCAT	247
181	CTCTCTCAACCTCAACCTCTTCCTTCGATCATCATCATCATCATCATCATCAT	240
248	CGCGGAGAAATGCGGAGTCCGATTCAGAACGCTGCTGCTTCACAGATTA	307
241	CGCGGAGAAATGCGGAGTCCGATTCAGAACGCTGCTGCTTCACAGATTA	300
308	TGAGAGTCTGCGAGATCATCCGATTCGCTCATTAATCATCTCTTGGCCCTAGG	367
301	CAGAAATCTAGCGAGATTCGATTCGCTCATTAATCATCTCTTGGCCCTAGG	360
368	ACGATTCGAGATTCGAGATTCGAGATTCGAGATTCGAGATTCGAGATTCGAG	427
361	TCTACAACTTCGAGATTCGAGATTCGAGATTCGAGATTCGAGATTCGAGAT	420
428	AACTTTAGCGGAGTGTGCTGCTTCCGCGGCGGCGGCTGCTGCTGCTGCTG	487
421	AACTTTAGCGGAGTGTGCTGCTTCCGCGGCGGCGGCTGCTGCTGCTGCTG	480
488	CGGACACACCGTTCAGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	547
481	CGGACACACCGTTCAGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	540
548	TCACTTAATTCGAGAGATGATGATTCGATTCGATTCGATTCGATTCGATTCG	607
541	TGAGCAACACATTCGAAATGAGA---TGAAGATGAGAATTAATTAATTCGCG	597
608	TCGACCGAGTCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC	667
598	TCGATGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	657
668	AGGAGCGTCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	727
658	TAAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC	717
728	ACCTTTACACCAATTCGAAATTCGAAATTCGAAATTCGAAATTCGAAATTCG	787
718	TCTTTTACACCAATTCGAAATTCGAAATTCGAAATTCGAAATTCGAAATTCG	777
788	ATGATTCATTCCT	847
778	GTTTATTCATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	837
848	AATCGGCGGAGATTCGAAAGAGTGTGATTCGAAAGAGTGTGATTCGAAAG	907
838	AATCATTCACGAGATTCGAAAG---CAAGACAAAGAGATTCGATTCGATTCG	894

OY	768	ATGATGAATACCTCTCTCCCTGCGCATGACGACCTAATATCTGGACGGCATCGCA	847
Db	778	GTTTAGTCATCTCTCATCTTTTGTAGCCATTGACGTTATATGTGCGTCGCATACGA	837
OY	848	AATCGTGCGAGATTACAAAGACGTGATGACGACAAACGCGCTTCTTTTAAATCCGA	907
Db	838	AATCATTCACAGCAGATGAAGAC--CAAGAACACAGATGCATTTTCTTGATCCGG	894
OY	908	TGCAATTTAAGCCCGCTTTGGATCCGCGGCTCCGGGGAACATCTTGGAACTGTCTAT	967
Db	895	TGATCTTAGGCCACAGATTAGATCCGCGGTCTCTAAATTAATCTTGGGAATCCTTAT	954
OY	968	CGTTTGGCATGGCGAMATCTCTGCGCGGAGTTTGGTTCGAGATGAAGGGTGTTCCGG	1027
Db	955	CGTACCGCGCTCCGAGAAATGGGGGGGAGAGCTGTGGGAGAGAAAGGGTGTTCTGG	1014
OY	1078	CAGCTGAGCGCATCGCGCGCGGAATAGACAAGACGACGCCACAAGATTTAGANA	1087
Db	1015	CAGCTGAGGTAATGCGCGCGAGATTAATAAAAGGATCAACGACAAAGAAATTTAGAA	1074
OY	1088	CTGTGAGAACTGCGCCTCTGAGATTCGCCAACCTTGCAGAAACTGTATTTCTCGGTGG	1147
Db	1075	CGGTGAGAAATGCTCCCGGAGATTGTGAACCGTTGCGAAGATATATTTTTCGGTGG	1134
OY	1148	CGGATGACAGCAGGCTTGATCTTTAAGCGCGGATTTTGGATGGGTAAAGCGGTGAAGC	1207
Db	1135	CAGGATCGAGCAACCTAGATCTTACGTCGCAATTTTGGATGGGGAAGCGGACAAAGC	1194
OY	1208	AAGAGATCTCTGCATTGATGAGAGAAAGTTTGCAGTTCGTTGCTTAACCGAGGAGAT	1267
Db	1195	AAGAAATTTTGTGCATTGATGGGAGAGAAATATCAATGACACTTTTAAAGCCAGGGATT	1254
OY	1268	CTGCGGAGAGATTGAGGTTGATGTCTTTGCCAAGAGAGAAATTCACAGCTTTTGATG	1327
Db	1255	TGCAAGGAGGATTGAGAGTTTGCTTCTTTCCTTAAGGACAAATGAGATGCTTTGCTG	1314
OY	1328	ATTATTTTGGGAGAGGAATAAGGCTTGATTAATCAATTAAATCATGATATATGAGTTGG	1387
Db	1315	CTTATTTTTCCTGGGAATTAATGTTAATTAATGATGTAATTAACATAATATGTGTGT	1374
OY	1388	ATGAATCTCTGTGTTCACTCTAATTTGTTAAACAATAATTTTTCATATGAACTTTT	1447
Db	1375	GTACAATTAATTAAGTGTGAGTAACGTGAAGATATACCTTATATATATTAATGATT	1434
OY	1448	TGAGCATATAAAAAAAAAAAAAAAAAAAAAATGAAAAA	1489
Db	1435	GGTCAAAATTAAGTTAAAGCCTCTGAAAAA	1476

RESULT 3	AB010708	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AB010708	1679 bp	mRNA					
			PLN					
			20-FEB-1999					
			Geniana triflora mRNA for Anthocyanin 5-aromatic acyltransferase,					
			complete cds.					
	AB010708							
	AB010708.1	GI:4185598						
			Anthocyanin 5-aromatic acyltransferase.					
			Geniana triflora petal cDNA to mRNA, clone:pgAT4.					
			Geniana triflora					
			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
			euphyllidophytes; Spermatophyta; Magnoliophyta; eudicotyledons;					
			Asteridae; Gentianeae; Gentianales; Gentianaceae; Geniana.					
		1 (bases 1 to 1679)						
	Fujiwara,H., Tanaka,Y. and Kusumi,T.							
	Direct Submission							
	Submitted (22-JAN-1998)							
			to the DDBJ/EMBL/GenBank databases.					
	Hirayuki Fujiwara, Suntary LTD., Inst. of Fundamental Research;							
	1-1-1, Mikiyamada, Shiamomo-cho, Osaka 618-0024, Japan							
	(E-mail:hirayuki.fujiwara@suntary.co.jp, Tel:+81-75-962-8807,							
	Fax:+81-75-962-8262)							
	2 (sites)							
	Fujiwara,H., Tanaka,Y., Yonekura-Sakakibara,K.,							
REFERENCE								
AUTHORS								
JOURNAL								

TITLE	CDNA cloning, gene expression and subcellular localization of anthocyanin 5-aromatic acyltransferase from <i>Gentiana triflora</i> Plant y. 16 (4), 421-431 (1998)
JOURNAL	
MEDLINE	
FEATURES	
	Fukuchi-Mizutani, M., Nakao, M., Fukui, Y., Yamaguchi, M., Ashikari, T. and Kusumi, T.
	Location/Qualifiers

source

C1

polya_site

```

/codon_start=1
/product="Anthocyanin 5-aromatic acyltransferase"
/protein_id="BAA74428.1"
/db_xref="gi:4185599"
/translation="MEQIQMKVLEKCGVMPSPSTMDVELSPVTFPDI PMYLHFNKQNS
SLLEDPPTPTHTFLDTVYIPMLKASLSITLKHVYPLSGNLMLPKSEEMKPFYSRBE
GDSITLLVAESDDVDYILKGIQLVDSNDLGLFVMPVRVITMDVYVPIPLVAQVNV
FPNGIAVNTLAHHSIADAKSPFMIINAMVITNFGEDADLISLNLKSPDRSIIRKL
YGLSEFTFEMQVDLEMFSEKSPFENKRAYVLSLAIEQILKLRKVNLRSQSEPT
IRVTFITMGQYVNTQMKSKDDVASEESNDNEKLEFSTACCRDLTPPCPEVF
GNCLASCAKATIKHELVGDKGLVAIVAIGAIEIKRLHNEGKVAIDAKTMLSSENGIP
SKRRIGTQSPKPDYGVGVRGCMKPAFDITSVYALIVYIQSRDEKVEICVSLP
KIHNDAPRAKIFEEGFCSL$"
1679
```

ORIGIN

Query Match	Score	DB	Length
12.7%	193	7	1679

Matches 684; Conservative 0; Mismatches 605; Indels 48; Gaps 8;

OY	46	ACGGGGGTGAGAGATCACTCCCGCTACCTTCTTCGACATGAGTGGCATTTTCAC	109
Db	69	ACAGATGTCGAGTTATCCCTACCGGTAACTCTTCGATATCCCTGGTTCAGTTAAT	128
OY	106	CCCATGCTTCAGCTTCTCTCTACGAACTCCCGTTCGAAACCCGCTTCTCGAACC	165
Db	129	AAGATGACGATCCCTCTCTTTTACGACTTTCGGTACCCAGAACATTTCTTGACACT	188
OY	166	GTCGTTCCGGAACATCAAACTCCTTATCTTAACCTCAAACTCTTCCCTTTCA	223
Db	189	GTTATCCCTTAATCTTAAAGCCCTTTTGCTCTCACTAAACACTAGTTCGCTTAGC	248
OY	226	TGCATCTCAATCTACCCCTATGCGGGAGAAATGCGGGATTCGAGTA-----TCAG	279
Db	249	GGAATTTTGTGATGCTCCATCAAAATCGGGCGAAATGCGGAAGTTTCAGTACTCCCGAT	308
OY	280	AACGGTGACTCGGTTCTTTCACGATTATVAGAGCTGTGCGAGA---TCATCCGATTC	338
Db	309	GAGGGCAGCTGATACCTTTGATGCTTCCGAGTCTGACCAAGATTTTGACTACCTTAA	368
OY	337	GCTCAATAATCTACTGCTTTGCCCTAGCGAGCATTAAGAATCTCCAGCTGCCCG	366
Db	369	GGTCAATCAAGTGTAGATTCCAAATGATTTGCATGGCCTTTATATGTTATGACACGGTT	428
OY	397	ATAGTCGAGGAATCTGATCGGAATTTGTTCAAGTTTACCGCTGCAAGTACTGT	456
Db	429	ATTAAGACCATGCAAGATATAAGTAGTATCCCGCTCGTAGCCGCAAGTAAACGTTTT	488
OY	457	CCCGGTGCGGGGGTGTGATCGGAATAACGAGCAACACACCGTTACGAGTGCCTCATG	516
Db	489	CCTAACCGTGGATAGCCGTGGCTCTGACGGCAATATCTCAATTTGACAGATGCTAAAAGT	548
OY	517	TTTGTGAGGTTTATGAAGTATGGGCTTCCATCACTAAATTCGAGAGATGATGAATTC	576
Db	549	TTTGTAAATGTTCAATCAATGCTTGGGGCTATTTAAACAATTTTGGGAAGAAGCGGCACTTG	608
OY	577	TTTGACGGAAGAGTGAAATGTTTGCCGGTTTTGACAGCATGCTGTGAATATATCCGCT	636

[illegible]

Query Match	Best Local Similarity	12.7%	Score 193;	DB 5;	Length 1703;
Matches 684;	Conservative 0;	Mismatches 605;	Indels 48;	Gaps 8;	
PN	JP 1997070290-A/1				
PD	18-MAR-1997				
PE	30-JAN-1996 JP 1996046534				
PR	17-FEB-1995 JP 95P 67159, 29-JUN-1995 JP 95P 196915				
PI	ASHIKARI TOSHITAKO, TANAKA YOSHIKAZO, FUJIWARA HIROYUKI, PI NAKAO MASHAHO,				
PI	PI FUKUI YUKO, YONEKURA KEIJO, MIZUTANI MASAKO, KUSUMI TAKAKI PC C12N15/09, A01H1/00, C07H21/04, C07K14/42, C12N9/10, (C12N9/10, PC C12N1:865),				
CC	(C12N9/10, C12N1:19);				
CC	strandedness: Double;				
CC	topology: linear;				
FH	key	Location/Qualifiers			
FT	source	1..1703			
FT		/organism='Gentiana triflora'			
FT		/variety='japonica'			
FT		/tissue_type='petal'			
FT		/clone='PGAT4'			
FT	CDS	6..1415			
FT	mat-peptide	9..1412			
FT		/product='acyltransferase precursor'			
FEATURES	source	Location/Qualifiers			
		1..1703			
		/organism='unidentified'			
		/db_xref='taxon:32644'			
BASE COUNT	512 a 350 c 359 g	482 t			
ORIGIN					

Db	609	TTGTCGCCGA-----ATCTTCCTTCACCTCTTGCATGAGATCGAATAATCAAGAATCTGAT	662
Qy	637	AAATTTGACACATATTTATGGAACAACGCCGAGAAGCTC-----CGTGGAAATCG	687
Db	663	GGCCCTAGAGGAACATTTTGGGAACGAATGCGAAMATGTTCTTGAATAGTCTCTAGATTT	722
Qy	688	CAGATCCATCTTTACCGACGAGATTCGAGATTCGAGTACCTACCTTTCCACCCATCTGAA	747
Db	723	GGAAACAAACCCCTCGATTCACAAAGGTACGACCTCAATATGCTCTCCCTTGCGAA	782
Qy	748	ATTAAAGAAATTTGAAG-----GGTTGATTCAGAGAAAGGCCCAATATGATTAATCTC	801
Db	783	ATCCAGAAAGCTTAAGGAACAAGTCTGATATTCGAGAGATTCGAAACCGACAATACGTGA	842
Qy	802	TCCTTCCTTCGTGCGGATTCGACGACTTATATCTGACCGGCATCCGCAATC-----	851
Db	843	ACGACGTTCAACAATGACGCTGTGATAGTATGACATGACATGCTGTCCTCAATCAAAAGATGAC	902
Qy	852	--GGTCGGAAGTTACAAAAGACGTGATGACAGCAACAACGCCCTTCTTTTAATTCGATC	909
Db	903	GTCGTATCTAGAGGATCATCGAACAGCAAGAAATAGTCAGATCTTCAGTTTACAGCG	962
Qy	910	GATTTAAGCCGCCCTTTGGATTCGCCGCCGCTCGGGGAACTACTCTGGAAATCGTCTATCG	969
Db	963	GATTGCGGAGACTTCTGACGCCGCCCGCTGCTCTACTACTTCTTGGCACTGTCTTCCG	1022
Qy	970	TTTTCGATGCGGAAGATCTTCGCGCGGAGTTTGCTCGAGATGAAGGGGTGTTTCGGGCA	1022
Db	1023	TCATGCGTTTGCAAAAGCAACACATTAAGAGATTAGTTGGGATTAAGGGCTTCTTGTTCA	108
Qy	1030	GCTGAGCGGATTCGCGCGCGGAATAGAGAAGAGACG---AGCCACAAGAAATTTCTGAA	108
Db	1083	GTTTCAGCTATTGGAGAAAGCCATTGAAAAGGTTGCACACGAAAGGCGTCTTGCA	114
Qy	1087	ACTGTGGAATCTGGCGCTGATTCGCGAAGCTTGCAGAAACGTGTTATTCCTGCTG	114
Db	1143	GATCGAAAAATTGTTATCTGGAATCTTAATGAATCCCTTCAAAAAGATTTCTCGGGATT	120
Qy	1147	CGCGGATCGAGACAGCTTGATCTTTACGCGCGGATTTGGATGGGCTAAGCGGTAAG	120
Db	1203	ACCGGATCGCCTAAGTTTCGATTCGTATGTGTAGATTTTGGATGGGAAAGCCTGCAAAA	126
Qy	1207	CAGAAGATCTGTGCAATGATGAGAGAAAGTTTACAGATGTCGTTGTGTAACCGAGGAT	126
Db	1263	TTTGACATTACCTCTGTTGATTAAGCAGAAT---GATTTATGTGATTCAGTCCAGGAT	131
Qy	1267	GCTGCCGAGAGATTGAGAGTTGATTTGTCTTTTGCCAAAGAGGAATTCAGACTTTGAT	132
Db	1320	TTTGAAAAAAGGTGAGAGATTGGATTCATTCATTCGCTTAAGATTCATATGATGATGCA	137
Qy	1327	GATTATTTTTCGGAGGG	1343
Db	1380	AAAATCTTTGAAGAAG	1396
RESULT	5		
LOCUS	E12757		
DEFINITION	Senecio cruentus mRNA for acyltransferase, partial cds.		
ACCESSION	E12757		
VERSION	E12757.1		
KEYWORDS	JP 1997070290-A/5.		
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 1508)		
AUTHORS	Ashikari,T., Tanaka,Y., Fujiwara,H., Nakao,M., Fukui,Y., Yonekura,K., Mizutani,M. and Kusumi,T.		
TITLE	GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY		
JOURNAL	Patent: JP 1997070290-A 5 18-MAR-1997; SUNTORY LTD		
COMMENT	OS Senecio cruentus PN JP 1997070290-A/5		

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
11.3%	Score 172.6;	DB 5;	Length 1508;			
48.2%	Pred. No. 2.7e-25;					
0;	Mismatches 739;	Indels 43;	Gaps			
7	ACCCCTCGAATCCCTCCGAGTGGCCGCGCTCCAGGACAGGTGGTGAAGCAGTCCAC	66				
3	AACATCTTGAACATGCGGATATGCGCCCGCGGGGACACATGGCGGATGCGTGA	62				
67	CCGCTACCTTCTTGCATGACGTGGCTGCATTTCCACCCCATGCTTACGCTTCTTC	126				
63	TCTCTACTTCTTTCGACATTAATGCTGCTACTCTTCCCTCGGTCACACATCTTTCTTC	122				
127	TACGACATCCCTCTTCCAAACCCGCTTCTCCGAAACCGTCTCCGAACTCAACAA	186				
123	TATGACTTTCACATCTTAAATCCCATTCATGACACATATGTTTCCAGGCTAAACAA	182				
187	TCTTATCTTAACCCCTCAACACTTCTTCCGCTTATGCAATCTAATCACTCA	246				
183	TCTTATGCGTACCTTCAACATTTTCCCGTTGGTAGTAATTTGATTTCTCT	242				
247	TCGCGG-----AGAAATGCCGAGTTCCGGTATCAGAACGCTGACTG	291				
243	AACACTGATGGTGGGTTTTTAAATAAAACAGAAATAAAACAGCTTGAAGTGATTC	302				
292	GTTTCTTACAGATTAATGAGTCTGCGAGATCAATCCGCAATCCGCTCATTAATAC	351				
303	GTTTGTGTTACTTTTGGCAAAATGTTGCTTGAATTAATTAATGACAGGAATCATCT	362				
352	TGCTTTGGCCCTAGCGAGATTAATGAATATCCAGTCCGCGCATAGTGAAGAACT	411				
363	CGAAATATGAAAACTTTATATCCACTTTACCTTATTTGGGAAATGCAATTAATAT	422				
412	GATCGGAAATGTTTCAAGTTTACCGCTGCAAGTCACTGTGTTCCCGGTGCGGGGT	471				
423	GATTCGCTCAGAGTCCACTTTTTCATTTCAATGAGAGAGTTTTCGCGGCTCGG	482				
472	TGCAATCGGAATTAACAGCAGCACACCGCTTACGATGCTCCATGCTTTGTAGG	531				
483	TCACATGAGAAATCAATCATATGATGCTTGGTGAAGGCTGACACCGGTTCAAT	542				
532	AAGATTTGGGCTTCATCAATTAATTCGAGAGATGATGAATTTGGACGAAAGT	591				
543	AAAGGTGACTTTCATTAATCTGCTGATGATGCTTTT-----AACGAA	596				
592	GAATGTTCCGCGTTTTCGACGATGCTGATGAATTAATCCGCTTAATTTGACATAT	651				

```

Db 597 GGATCTCCACCGGTTTGTATGATTGATT---AACATCCACATTTAGATGAAATAAG 653
Oy 652 TTATGGAACACGCGAGAAAGCTCCGTTGATGACGACATCCATCTTACCAGGAT 711
Db 654 TTGAGACATACAAAGCTCGAAAGTTTATTAACCTTCGAGCTTGTGTCTCCACTGAT 713
Oy 712 CGGATTCGAGTACCTTACCTTTTCAACCCATCTGAAATTAAGAAATTAAGAGTTGAT 771
Db 714 AAAGTTCCGTCACGTTTGTGTGAACCGCACTAATATCACTCTCTAAGAAAGAGTTC 773
Oy 772 CAGAGAAAAAGCCCAAAATGATTAATCTCTTCTTCCTGCGCATGCGACCTATATC 831
Db 774 TTAAACCAAGTCCCAAACTGGAGTACATGATCTATTTACGTAACCTGTGTGTATATA 833
Oy 832 TGGACCGCATCGCAAAATCGGTGGAGATACAAAGA---CGTGATGACGACCAAGCC 888
Db 834 TGGATTCGATGACCAAAATCACTCTAAATAAGAGAAAGAGGCGCAACAGATTA 893
Oy 889 GCTTTCTTTTAAATTCGATCGATTTAAAGCCGCTTTGATCCGCGCTCCGGGGAAC 948
Db 894 GAACAGTTCATATCACCATTGATGTGATCTGATCTGATCCACCAATTCACACAGCC 953
Oy 949 TACTTCGGAACCTGCTATCGTTTCGATGCGGAATCTCGCGCGGATTTGTCGGA 1008
Db 954 TACTTTGTTAACTGTGTGACACCATGTGTCCGACCTTAAAAAATGTGCTTTGACTAGC 1013
Oy 1009 GATGAAGGGGTGCTTCGGGACACCTGAGCGATCGCGGGAATAGAGAGAGGACGAC 1068
Db 1014 GAAATGGGTATGACACTGTGTCTAAAGTAATTTGAGAGATCTATATGCAAAATGATATAT 1073
Oy 1069 GACAAAGAAATTTAGAAACTGTGAGAACTGGCCGCTGAGATTCGGGAACCTTGCAA 1128
Db 1074 AATAAGGACGGAATCTGAAAGATGCCGAGATGGACAAGAACCTTTCATGATCCCGCT 1133
Oy 1129 AACTGTTATTTCTCGTGGCGGAGATCGACAGCGCTTGAICTTTACGCGCGGATTTTGA 1188
Db 1134 A---GGAAGATTTGTGTGTGTGTACACCTAAGCTCAACTTGTGACTTTGATTTTGGG 1190
Oy 1189 TGGGTAAAGCGGTGAACAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1248
Db 1191 TGGGGAAAGCGGATTAAGTATGAGACTGTTCAATAGACTATATATAGCT---CGATTTCT 1247
Oy 1249 TTGTGTAACCGAGGATGCTGCGGAGAGATTTGAGATTTGATGATTTGCGCAAGAG 1308
Db 1248 ATAAATGCAACCAAAACATCTGACACAGATCTTGAATTTGATGATGATGATGATG 1307
Oy 1309 GAATTCGAAGCTTTGATGATTTTTCGAGAGGGAATTAAGGTTGATTAATCATTTTA- 1367
Db 1308 CAATGAGGCGGTTTCTAGCATCTTTGATGATGATGATGATGATGATGATGATGATG 1367
Oy 1368 -----ATCAGTATTTAGAGATTTGATGATGATGATGATGATGATGATGATGATG 1418
Db 1368 ATCATGTCGCCCTTTTGTGTGATCAAGTTCTGTCTTTTATGATGATGATGATGATG 1427
Oy 1419 AACATATATTTTTCATGATGATTTTTCATGATGATGATGATGATGATGATGATGATG 1478
Db 1428 TATCTTAACTATACCTTTGATGATGATGATGATGATGATGATGATGATGATGATG 1487
Oy 1479 AAATGAAAAA 1489
Db 1488 AAAAAAAAAA 1498

```

```

RESULT 6
LOCUS E12754 1622 bp DNA PAT 24-JUN-1998
DEFINITION Gentianatriflora mRNA for acyltransferase, complete cds.
ACCESSION E12754
VERSION E12754.1 GI:3251586
KEYWORDS JP 1997070290-A/2.
SOURCE unidentifed.
ORGANISM unclassified.

```

```

REFERENCE 1 (bases 1 to 1622)
AUTHORS Ashikari,T., Tanaka,Y., Fujiwara,H., Nakao,M., Fukui,Y.,
TITLE Yonekura,K., Mizutani,M. and Kusumi,T.
GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY
JOURNAL Patent: JP 1997070290-A 2 18-MAR-1997;
SUNTORY LTD
COMMENT OS Gentianatriflora
PN JP 1997070290-A/2
PD 18-MAR-1997
PE 30-JAN-1996 JP 1996046534
PR 17-FEB-1995 JP 95P 67159, 29-JUN-1995 JP 95P 196915 PI
ASHIKARI TOSHIHIKO, TANAKA YOSHIKAZU, FUJIWARA HIROYUKI, PI NAKAO
MASAHITO,
PI FUKUI YUKO, YONEKURA KEIYO, MIZUTANI MASAKO, KUSUMI TAKAARI PC
C12N15/09,A01H1/00,C07H21/04,C07K14/42,C12N9/10,(C12N9/10, PC
C12R1:865),
PC C12N9/10,C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH key Location/Qualifiers
FH source 1..1622
FH FT /organism='Gentiana triflora' FT
FT /variety='japonica',
FT /issue_type='petal',
FT /clone='pGAT106'
FT CDS 35..1474
FT /product='acyltransferase'.
FEATURES
source Location/Qualifiers
1..1622
/db_xref='taxon:32644'
BASE COUNT 478 a 345 c 322 g 477 t
ORIGIN
Query Match 10.18; Score 153.8; DB: 5; Length 1622;
Best Local Similarity 49.48; Pred. No. 1.8e-21;
Matches 692; Conservative 0; Mismatches 652; Indels 57; Gaps 9;
Oy 4 ACCACCTCTCGAATCTCCGAGTGGCGGCGCTCCAGCAGCGTGTGACGATCA 63
Db 56 ATCAAGTTCCTGAGAAATGCGGTGTGCGCACACCGAGCGCGCGAGTTTACA 115
Oy 64 CTCGCCCTACCTTCTTGACATGACGTGCTGCAATTTCCACCCTATCTTCTC 123
Db 116 GTCCACTGTGTTTGTGACATGAGATGTTGATCTGTATGCAACACCATCTGCAT 175
Oy 124 TTCTAGCAACTCCCTGTTCACAAACCGCTTCCGTAACCGTGTCCGAAACTCAA 183
Db 176 TTCTAGCAATTCGCGCATCTTGTCCCAACTCTA---AATTTATCTTTCATTTAA 232
Oy 184 CAATCTATCTCAACCTCAACACATCTTCCCTTCATGCAATCTATACCT 243
Db 223 TCGTCCCTTCCCTGTTCACAAACATCTTCCGTTAGCGGGAATTTGATTTGCGCG 292
Oy 244 CTA---TCGCGGAGAAAATGCGGAGTTCGGGTATCAAGACGTTGATCTTTTC 300
Db 293 GTAGATTCCTCGATAGATGCGGAGTTGGCTTAACAAGAGGGACTCGGTTCTTTA 352
Oy 301 ACGATTTAGAGTCTGTGAGATCATTCGATTCGCGCTCTAATATCTACTGTTGCC 360
Db 353 ACAATTCGAGATGCGAGATGATTTGATTTATCTCGCGGAGATCATCAGAGGATTC 412
Oy 413 TATTAATTCACAGATTTGATTCGCGAGCTGCCAAGACGATTTTAACCTCGGAGAGAA 472
Db 421 TTGTTCAAGTTTAAAGGATCTCAGCTGCGCGGAGATGTCGAGATTCGAGAA 480
Oy 473 GTATTACCACTTTTGTCTTACAGTGAAGTGTCTTCCAAACCGGTTATATGATGCA 532
Db 481 ATTAAGACGACACACCGTTAGCATGCTTCATGTTGAGGCTTTATGAAGATGG 540

```

```

Db      533  CGCAATCTTCATCAAGCTTCTGGTGTATGCGAGTTCCTTTCTTCGCAATTTATATATATATG 592
Oy      541  GCTTCATCACTAAATTCGGAGGATGATGATTTCTTGAGCGAAGAGTGAATTTTG 600
Db      593  GTTTTGTTGACAAATCCATGAGAGATTC---ATTAAAGTCTCTTCCACTTCTCTCTGA 649
Oy      601  CCGGTTTTCAGCAGATCGCTGCTGATTAATCCGCTTAATTTGACACATATTTATGAAAC 660
Db      650  CCATGTAGACAGATCTGTGGTGAAGATCCATTTCAATTCGTGGAAGAAATCTACAT 709
Oy      661  -----AACGCCAAGAACGTCCTTGGAATC-----GCAGATCCATCTTTACCGAGC 708
Db      710  GAAGAAACCTGCTCAAACTCTAGGCGACACCTACTGTTCTTAATCCAGCAATTTCTAAA 769
Oy      709  GATCGGATCTGAGCTACCTACCTTTACCCCATCTGAATTAAGAAATTAAGAGGTTTG 768
Db      770  GATGAAGTTTGAGCCACCTTCATCCACACCTTATGATATCATGAGCTCAAGAAATTC 829
Oy      769  ATTCAAGAGAAAGCCCAATGTA-----GTTAATCTCTTCTCTCTGTC 813
Db      830  ATTGCTCAAAATATGCAACTTAACCGGATGATTAATTAATTAATCTGCACTTTCAGC 889
Oy      814  GCGATCGCAGCTTAATATCTGAGACCGGATGCGCAAT-----CGTCCGAGATTAAC 864
Db      890  GTGACATCTGCACTGATGAGCATGCTGTGCAATCATATGACACCGCTGTAAGAGAG 949
Oy      865  AAGAGCTGGATGAGCAACAACGCGCTTCTTTAATCCGATGATTAAGGCGCGCT 924
Db      950  AAGGTGAAGAGATTAACATGACGCAAACTTAATGCTTCTTCACTGCGCAACAGCT 1009
Oy      925  TTGGATCCGCGCGCTCCGGGGAATCTTCGAAACTGTCTATCTGTTGCGATGGCGAAG 984
Db      1010  TTTTGCTCCCGCGATCTCTCAAAATTAATTTGAAATTTGATGATGCGCTGTATGTTGG 1069
Oy      985  ATCTCTCGCGCGGATTTGCTCGAGATGAAGGGCTTTTCGGGACCTGAGCGGATCGCG 1044
Db      1070  TCGACTCATGAGCAACCTTTGAGGAATGAAGGGTTGCTGCTGACTGCAACCGCATCGGA 1129
Oy      1045  GCGGAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1101
Db      1130  GATGCTATCATTAAGAGGTTTACATGACTACGAGAGAAATTTCTGAGAGAGATTTGATAT 1189
Oy      1102  CCGTCTGAGATTCGCGAAGCTTGCAGAACT-----GTTATTTCTCGGTGGCGGA 1152
Db      1190  CGGCCCGCATCAACATCTGCGGACCAAGTCCAGCTCATTTATGTCGTTGATCCGGA 1249
Oy      1153  TCGAGCAGGCTGATCTTTACGCGCGGATTTTGATGGGATGAAGCGGTGAAGCAAGAG 1212
Db      1250  CAACGCAATGTGCATATTTTGTGATGAGATTTTGTGGGAAAGCTTGAAAGCATGAA 1309
Oy      1213  ATACTGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1272
Db      1310  TCTGTTCACTAATATCTCTTCGCAACATTAATTTATCTCTGCTGCTCAGAGAGATTAAA 1369
Oy      1273  GAGAGATTGAGAGTTGATTTGCTTTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1332
Db      1370  GAGGCACTTGAGGTTGGCAATTTCTTGGCCAAAGATAGATGAGAGCATTTGGCACAT 1429
Oy      1333  TTTGCGGAGGAGATTAAGGCT 1353
Db      1430  TTTACGAATTTCAATCAATAGT 1450

```

```

RESULT 7
AB026494 1622 bp mRNA PLN 04-APR-2000
LOCUS AB026494 1622 bp mRNA PLN 04-APR-2000
DEFINITION Gentiana triflora GAT106 mRNA for acyltransferase homolog, complete
cgs.
ACCESSION AB026494.1 GI:7415596
VERSION AB026494.1
KEYWORDS acyltransferase homolog.
SOURCE Gentiana triflora cDNA to mRNA.

```

```

ORGANISM Gentiana triflora
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; eustersids I;
Gentianales; Gentianaceae; Gentiana.
REFERENCE 1 (sites)
AUTHORS Yonekura-Sakakibara,K., Tanaka,Y., Fukuchi-Mizutani,M.,
Fujiwara,H., Fukui,Y., Toshihiko,A., Yamaguchi,M. and Kusumi,T.
TITLE Molecular cloning and biochemical characterization of
hydroxycinnamoyl-CoA:anthocyanin
3-O-glucoside-6-O-hydroxycinnamoyltransferase from Perilla
frutescens and diverse plant acyltransferase homologs
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 1622)
AUTHORS Tanaka,Y. and Yonekura-Sakakibara,K.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1999) to the DDBJ/EMBL/GenBank databases.
Yoshikazu Tanaka, Suntory Ltd., Institute for Fundamental Research;
Wakayama-dai 1-1-1, Shimamoto, Osaka 618-8503, Japan
(E-mail:yoshikazu.Tanaka@suntory.co.jp, Tel:81-75-962-8807,
Fax:81-75-962-8262)
FEATURES
source location/Qualifiers
1..1622
/organism="Gentiana triflora"
/db_xref="taxon:55190"
35..1474
/gene="GAT106"
35..1474
/gene="GAT106"
/codon_start=1
/product="acyltransferase homolog"
/protein_id="BA03452.1"
/db_xref="GI:7415597"
/translation="MAGNSEDIKTVLEKCRVAPPDPAEFYPLSPFDMRLISDAEH
HIEYFRPHPCNSKEFISISKSLSTLVKLFHPLAGLIMPVDSDDMPRLRYKGD
SVSLTIAESSMDEPLIADRDHDSKRFNDLIPQLEPIVTSDEVLPLFAIQYFESN
TEICGRNDHOVLDASSFLEHNLKLVAVDSNGSLFLPLUSLPMYDRSVDDPRH
IRKILYNERKILKSQGTPTVPLPALSKVDEKVAFTFLHPIDILMLKLFISNRLTGS
SNVNLSTFTVLSALIMTCLSKSLDVLVEKEEDKHAANLCAFIKRGORFAPLPQNY
FGNCIVPCMGSTHEDQLVNGEGLSVAALAIKRIKRLHDEGLLRGDMISPRSTSA
APRSLTIVVGSAGRNVDHPDADPFMGLEKHESVSTNPSATLILISRRRKALGL
GISLPKRNMDAPATIEFTNPLINSHRSL"
BASE COUNT 478 a 345 c 322 g 477 t
ORIGIN
Query Match 10.1%; Score 153.8; DB 7; Length 1622;
Best Local Similarity 49.4%; Pred. No. 1.8e-21;
Matches 692; Conservative 0; Mismatches 652; Indels 57; Gaps 9;
Oy 4 ACCACCCCTCTCGAATCTCCGAGTGGGCGCGCTCCAGAGCAGGCTGCTGAGCATGA 63
Db 56 ATCAAAAGTTCTTGAGAAAGCGGTGTGGCGCACACGAGCGCGTCCGCGATTACA 115
Oy 64 CTCGCCCTACCTCTTCGACATGACGTGCGATTTCCACCCCATGCTTCACTTTC 123
Db 116 GTCCCACTCTCGTTTTCGACATGCGATGTTGATCTGATGACAAACACATCTGCAT 175
Oy 124 TTTACGACTGCCCTGTCCAAACCGCGCTTCTCGAAACGCTGTCCGAAACGTCAA 183
Db 176 TTCTACGATTCGCGCATCTGTTGCCAACTCTA---AATTTATCTATTCATCATTA 232
Oy 184 CAATCTTATCTTAACCTCAACACTTCTCCCTTTCAATGCAATTAATCTACCT 243
Db 233 TCGTCCCTTTCCTTCTTCTCAAAACACTTCTTCGTTAGCCGGGAAATTTGATTGGCG 292
Oy 244 CTA---TCGCCGAGAAATTCGCCGAGTTCCGGTATCAGAACGATGATCGGTTTCTTC 300
Db 293 GTGATTTCTCTCCGATGAAATGCGCGAGTTGCTTACAAAGAGGAGATCCGTTTCTTTA 352
Oy 301 AGCATTAATGAGATCTGTGGAGATCATCGCATTCGCGTCAATAATTAATGATGCTTGC 360
Db 353 ACAATTGCAAGATGAGCATGATTTTGAATTAATCTGCGCGGAGATCATCAGAGGATCT 412

```


Oy	361	CCTACGACGATTGTGAAGATCTCCAGCTCCGCCGATAGTCGAGAGATCTGATCGAAA	420
Db	413	TATTAATTTCAACGATTTGATTTCCGACGTGCCAGAAACGATTTGTAACCTCCGGGACGAA	472
Oy	421	TTGTTTCAGATTTTAGCCGTGCAGAGAGACGTCGTTTCCCGGTCCGGGGTGTGCATCGA	480
Db	473	GTATTACCACTTTTGTCTTTACAGGTGCACGGTGTCTCTCAACACCGGATATGCGATTGGA	532
Oy	481	ATAACGACGCCACACCCGTTAGCGATGCTCCATCGTTTGTAGGGTTTATGAAGAGTTGG	540
Db	533	CGCATCTTCATCAAGTTCTTGGTGTATGCCAATCTTTTCTTGATCTTTTAAATAATATGG	592
Oy	541	GCTTCGATCACTAAATTCGGAGAGATGATGAAATCTTTGGACGGAAAAGTGAATGTTGG	600
Db	593	GTTTTGGTTGACAAATCCATGAGATTC--ATTAAATCTCTTCCACTTTCTCTCTGA	649
Oy	601	CCGGTTTTTCAGACGATTCGCTCGTGAAATTTATCCGCTTAATTTGACACATATTTATGTAAC	660
Db	650	CCATGTACACAGATCTGTGGTGCAAGATCCATTTCAATTCGTGCMAAAATCTCAAT	709
Oy	661	-----AACGCGCAGAAAGCTCCGTTGGATC-----GCAGCATCCATCTTACGAGAC	708
Db	710	GAAAGAAAATCGTCAAAATCTCAGGSCAACCTACTCTGTTCAATTCAGCAATTTCTAAA	769
Oy	709	GATCGGATTCGAGCTACCTACTCTTTTCACCCAACTGAATTAAGAAATTTGAAGGTTTGG	768
Db	770	GATGAACTTCGACCACTCACTCTACCTCTACACCTATGATCATGTGAAGCTCAAGAAATTC	829
Oy	769	ATTACAGAAAAGCCCCAAATGTA-----GTATATCTCTCTCTCTCTCGTC	813
Db	830	ATTTCGTCAAAAATATGCAACTTAAACGCTAGTGTATTAATATCTGTCAAACTTTCACG	889
Oy	814	GCGATGCGACGCTTATATCTGGACCGCGCATGCCCAAT-----CGTCCGAGATATAC	864
Db	890	GTGACATCTGCACCTATCTGGACATGCTGTGCAGAACTTATACACACCTCTGTAGAGAG	949
Oy	865	AAAGACGTGATGACGACAAACGGCTTTCTTTTAATTCGATGCAATTTAAGCGCGGCT	924
Db	950	AAGGTGGAAGAGATTAACATGACGAAACTTATGTGCTTTCATCACTGCCGCAACGT	1009
Oy	925	TTGGATCCGCGCGCTCCGGGGAATCTACTGCGAAACTGTCTATCGTTTGCATGGCGAAG	984
Db	1010	TTTTGCTCCGCCATPACCTTCAAAATTTACTTTGGAAATTTGATAGTACCTGTATGTGTGGA	1069
Oy	985	ATCTCGCGCGGATTTTGGTCGAGATGAAGGGGTTCGCGGACGTGAGCGGATGGCG	1044
Db	1070	TCGACTCATGACCAACTTTGAGAAATGAAGGGTTGTGCTAGCTGCACACCCCATGGGA	1129
Oy	1045	GCGGAAATAGAGAGAGAGAGAGCGACGAAAGAA---GATCTTGAAACTGTGAGAACTGC	1101
Db	1130	GATGCTATTCATTAAGAGCTTACATGACTCGAAGAAATTCGAGAGAGATTTGGATATTCG	1189
Oy	1102	CGGTTTGAGATTCGCGAAGCCTTCGAAAAC-----GTTATTTTGTGGTGGCGGA	1152
Db	1190	CCGGCCCGCATCAACATCTGGCGCACCAAGGTGACGCTCATTTATGTGTTGGATCCGCA	1249
Oy	1153	TCGAGCAGGCTGTATCTTTTCGCGCGCGGAAATTTGATAGGGGTGAAGCAAGAC	1212
Db	1250	CAAGCAATGTGCATGATTTTGTATGCAAGATTTGTGTGGGAAGACTTTGAAAGCATGAA	1309
Oy	1213	ATACTGTGCATTGATGGAAGAGATTTACGATGTGTTGTGTAAACCGAGGATGCTGGC	1272
Db	1310	TCGTGTTCAACTATCTTCGCAACACTAATTTTGATCTGTGGTCCAAATAATTTAAA	1369
Oy	1273	GGAGGATTTGAGGTTTGATTTGTTCTTTGGCAAAAGAGAAATTCGAAGCTTTTGATGATAT	1332
Db	1370	GGACACTTGAGTTTGCGCATTTCTTTGCTTAAGATTAAGATGACGCAATTTGCCACATT	1429
Oy	1333	TTTTCGGAGGAAATAAGGCT	1353
Db	1430	TTTACGAAATTTCAATCAATAGT	1450

RESULT	8
AB016892	AB016892 66237 bp DNA PLN 20-NOV-1999
LOCUS	Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone: MXF12,
DEFINITION	complete sequence.
ACCESSION	AB016892
VERSION	AB016892.1 GI:3449333
KEYWORDS	HTG.
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_1lb:Mitsui P1 Clone:MXF12.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (sites)
AUTHORS	Nakamura, Y.
TITLE	Structural Analysis of Arabidopsis thaliana Chromosome 5. VIII
JOURNAL	Unpublished (1998)
REFERENCE	2 (bases 1 to 66237)
AUTHORS	Nakamura, Y.
TITLE	Direct Submission
JOURNAL	Submitted (18-AUG-1998) to the DDBJ/EMBL/Genbank databases. Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2, 1532-3, Yana, Kisarazu, Chiba 292, Japan (E-mail:y.nakam@kazusa.or.jp, Tel:+81-438-52-3935, Fax:+81-438-52-3934) Location/Qualifiers 1. 66237 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /chromosome="5" /clone="MXF12" /clone_1lb="Mitsui P1"
BASE COUNT	21122 a 11849 c 12039 g 21227 t
ORIGIN	
Query Match	7.0%; Score 106.4; DB 7; Length 66237;
Best Local Similarity	46.5%; Pred. No. 1,le-11;
Matches 664; Conservative	0; Mismatches 716; Indels 48; Gaps 8;
Oy	45 CAGGTGGCGTGCAGCATCTCCCGCTCACCTTTTCGACATGACGTGGCTGCATTGCCA 104
Db	34419 CTGTGGCGCTCACTCACTCCCGCTCACTTTTGAGACTTCCTGCTCAAAACACA 34478
Oy	105 CCCATGCTTCACGCTCTCTTTCAGAACCTCCCCTGTTCCAAAACCCTTCTCTGAAC 164
Db	34479 AGCGTGCMAACGATGATCTTCACAAATCACCAGCATTAATTCGTCTTATTCGAC 34538
Oy	165 CG----TCGTTCCGAATAACAATCTTATTCCTTAACCTCAAAACACTTTCGCCCT 221
Db	34539 AGTCATCTGTCCTCCCACTCAAAATCCCTCTCTCTCTCATCTCTCTCACTACCTCCACT 34598
Oy	222 TTCATGATCTAAATCTTACCTCTCTATGCGCGGAGAATAATGCCGAGTTCCGGATCAGAA 281
Db	34599 CGCGGCCATTCATCT---GGAAACCATATATCCAAAAACGAAATTCGTACACCCA 34655
Oy	282 CGGTGACTCGTTTTCTTCACGATTATGAGTCTGTGGAAGTCATCCGATTCGGCTCA 341
Db	34656 AAGACGCGCTTTCATCTACCGTAGTGAAGCAACATCGGATTTCTCCCTTTAACCG 34715
Oy	342 TAAATACATCTGCTTTGGCCCTAGGAGAGATTATGAGATCTCCAGTCCGCGCATAGT 401
Db	34716 AAAAATACCTTCTCTCCACCGAGTTACACCCATTAATCTCCGAGTTACAAAATCTCGA 34775
Oy	402 CGAGAAATCTGATCGGAATTTGTTCAAGTTTATAGCGTGAAGTGAOTCGTTTCCCG 461
Db	34776 CGACTCACACAC-----AGGTGTCTCGTTTCAATCATCGTTATTTTCCAAA 34820
Oy	462 TCAGCGGTGTGCATTAACGACGACACACCGCTTACGAGATCTCATCTTTGT 521

Db 34821 CCAAGGTTTTGCATCGCGGTAAACACACACATGCCGTTTCATGGGAAAAAGCAAC 34880
QY 522 AGGTTTATGAAAGATTGGGCTTCATCACTAAATTCGAGAGATGATGATTTCTTGA 591
Db 34881 CAGCTTTCCAAACTTTGGGCTCAGTTATGTAA-----CATCAAACTCTTC 34928
QY 582 CGAAAAAGGTAATGTTTGGCGTTTTCGACCGATCGTCGTGAATTAATCCGCTTAAT 641
Db 34929 TCTACCGAGATGATCAATTCGTTTACGATCGTACGTCATAAAGTCCACCGAGAT 34988
QY 642 GGCACATATTTATGAAACACCGGAGAAAGCTCCGTTGGATCCGACATCATCTT 701
Db 34989 TGTACTTAAGTCTTGAATTTGGACCTCTATTCATTAACCAAAAGCTTGAAGTTGT 35048
QY 702 ACCGAGATTCGATTCGACCTACCTCTTCCACCAATCTGAATTA--AGAAATTC 759
Db 35049 ACCGAGACCGAGATAGAAATCCAGATCGTTCATACATTCGAATCTACTGTGAAA 35108
QY 760 AAGGTTTATTCAGAAAAAGCCCAAAATGATTAATCTCTCTCTTCCTGTCGCGATC 819
Db 35109 CATCGAAAAAGCTTCGAGATTAACCTCAAGAGAGATCTCTCTGTTTCCGTTAGATT 35168
QY 820 G-----CAGCTTATATCGAGCGGACGCAATCGGTCGAGATTAACAAGCT 872
Db 35169 GTCAACGTTTGTGATTAAGTTTCTGTAACGTTTACGTCCTTAATAGATCCGAGCGCA 35228
QY 873 GGATGACGACAAAGCGCTTCTTTTAAATCCGATCATTAAGGCGCGCTTG--GA 929
Db 35229 TGATTCGAATAGACGCGGTGGGATATGATTCGCGGTGATGTCGAAGGCTTATCATGAT 35288
QY 930 TCCGCGCGCTCCGCGGAACTACTTCGAAACCTGTCTATCGTTTGCATGCGAAGATCCT 989
Db 35289 TCCACCGATCCCGTACTTATTCGAAACCTGTTTACTCTGCTGTGAATAATCCCGCT 35348
QY 990 GCGGCGGATTTGCTGAGATGAAGGGGTTCGCGGACGTCGAGCCATCCGCGCGCA 1049
Db 35349 AGATCGCGGAGATTTCTTGGTGTAACAAGGTTCTGTTGTCGCGAGATTTGATGTA 35408
QY 1050 AATGAGAAAGAGACGACGACAGAAAGATTCAGAACTGTGAGAACTGCGCGCTCTGA 1109
Db 35409 TTGCGTTGAGGAATTTGATTCAAATGTGGCTTGGAAAGATACCGGACCTTTTGGAAACGTA 35468
QY 1110 GATTCGCGAAGCTTCGCAAACTGTTATTTCTCGGTCGCGGATGACGACGCTTGAATCT 1169
Db 35469 TGAAGAAAGTCTCTGTGACTCAAAATTTGTGTCGTTGCGCGGTCAACCGGTTTGTAT 35528
QY 1170 TTACGCGCGGATTTTGGATGGGTAAAGCGGTGAAGCAAGATATCGATTTATG 1229
Db 35529 ATACGCGTTGGATTTGGGTGGGAAACCTTTTAAGTCGTTGTGTGTCGATTCATTC- 35587
QY 1230 AGAAGATTACGATGCTGTGTAAACCGAGGATGCTGCCGAGAGATTTGAGGTTGG 1289
Db 35588 --AGAAAGGAAAGATATCAATTCGAGAGATGAGATGAAAGTGGGCTTGAATTTG 35645
QY 1290 ATTGCTTTTCCAAAGGAGATTTGCAAGCTTTGATGATTAATTTTCCGAGAGGATTA 1349
Db 35646 CTCTCTGCTCAAGAAACAAAGAAATGATTTGATTTGATTTCTCTACACAAGGGGATTA 35705
QY 1350 GGGTGTATTAATCATTTAATCATGATTAATGAGTTGATGAAATCTCTGTTGATCTC 1409
Db 35706 ATGATTAATGACAAATTAATTAATTAATTCAGTGTCTCTCTTGTAGCTTGCATCGA 35765
QY 1410 TATGTTTAAACAATAATTTTTCATTCGAACTTTTTCAGTCAATA 1457
Db 35766 TTTGGTTCTGAACATTAATGTAACCTTGTGATTTGATGTTCAATA 35813

RESULT 9
AC002560 101176 bp DNA PLN 19-JUL-2000
LOCUS AC002560 101176 bp DNA PLN 19-JUL-2000
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F21B7 from chromosome
I, complete sequence.
ACCESSION AC002560

VERSION AC002560.2 GI:9211167
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS Khan, S., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Kim, C.,
Shim, P., Altafi, H., Bel, Q., Chin, C., Chlou, J., Choi, E., Conn, L.,
Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B.,
Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharsky, N.,
Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A.,
Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Federspiel, N. A.,
Theologos, A. and Ecker, J. R.
Genomic sequence for Arabidopsis thaliana BAC F21B7 from chromosome
I
Unpublished
2 (bases 1 to 101176)
Ecker, J. R.
Direct Submission
Submitted (25-SEP-1997) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 101176)
Ecker, J. R.
Direct Submission
Submitted (13-NOV-1997) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 101176)
Ecker, J. R.
Direct Submission
Submitted (26-JAN-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
5 (bases 1 to 101176)
Cheuk, R., Shin, P., Brooks, S., Buehler, E., Chao, Q.,
Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bel, B., Chin, C.,
Chlou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N.,
Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J.,
Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H.,
Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M.,
Yu, G., Davis, R., Federspiel, N., Theologos, A. and Ecker, J.
Direct Submission
Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
6 (bases 1 to 101176)
Ecker, J. R.
Direct Submission
Submitted (15-JUL-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
7 (bases 1 to 101176)
Cheuk, R., Shin, P., Brooks, S., Buehler, E., Chao, Q.,
Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bel, B., Chin, C.,
Chlou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N.,
Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J.,
Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H.,
Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M.,
Yu, G., Davis, R., Federspiel, N., Theologos, A. and Ecker, J.
Direct Submission
Submitted (19-JUL-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Jul 15, 2000 this sequence version replaced gi:2618677.
Location/Qualifiers
1. 101176
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F21B7"

CDS
join(736..1038,1276..1713,1874..2119,2205..3032,
3117..3274,3483..3610,3725..4134,4313..4435)
/note="unknown protein: similar to ESTs dbj|AV529515.1,
dbj|AV523761.1, dbj|AV54950.1, dbj|AV523761.1,
dbj|AV523761.1, dbj|A1996703.1, and dbj|AV546743.1"
/codon_start=1
/evidence=not_experimental
/product="F21B7.1"
/protein_id="AAF86508.1"
/db_xref="GI:9280639"
/translation="MQPLPARSGDHEGFNNSHPLILVAGDETNCTGLGHSFQNSGL
ARDGSDSKAGDAINPTVRYSLSHSYVYVLRSSVCMIRCSRYVAAGLANOI
YCVADLTLENKSVLYTPVPOVROGTTPVNGVGPMAVGPMLVAASKSMNTGR
LSRPTSPSPSLSPSSSGSSPMARYAMESKOLANGLINGDGYKTLSCYCDML
PDGSTPASPNAIMKYGSGSDAENAGVAVKDLVSGALVSOPKARHSPIALCFDP
SGTLVYASVGNNTINVOIMPSRSHNAGDLSYEMSHVLFKLRHCTISAIYODI
CPSOOSQVAIISSKGTCHITPYNSSGDAAPCEGGEPTPLPASPSPMPFTQSLSS
NOQSLSPYVALSVSRKYSFGMLNTVSATTAATGKVPSPGAVAAVFKKSVTH
DIQLNSRTNALLEHLYTPSGHVVOHELPSYCTESPENGLRVKTSHVQVEDDLRV
KVEPIOMDVEKRRSDMLEERLPKSTIKOYDLETVSNHLTSHEDACSLDMNHS
EDKYLKSCSEKPEPSHCHLSNFEYKVTSGMLPMONKTSIFHWDSRDSSTGGEF
EIEKPAHELEIKOKKLPVDPHSTKATLEDPRSMKCTHSATGSHQVCKTODI
INCHSKRGSIBESSEBSTRKQEMENLHDSHMSIKSSLPYPTVNGIKYKEIKNN
ANGMEKEPVTAKLSTLKEKRTINGFTTPILDSVNEQMLSTGKPMGFGFALHEHC
KAVADPKKEHLKKKLDEVTNHHNLVNNNNTEKLOGDKVNSQVLANFAGVCILYKDFL
RVGNMFSVPFLLYKILRK"
complement(4838..6223)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F21B7.2"
/protein_id="AAF86541.1"
/db_xref="GI:9280672"
/translation="MASCIOELHFTLHIPTINOQFLVHPSPSPRANOSPHHSIYS
NLDDIGARVTPSYVEYESTNNRESFPLKRLQDLSVLYPIPLSKLKEVEKGL
EYFGEBOGVWYNSANSSMDLADGLVVPAMLPILFRNPGGEAYLILMPLRLN
VTPPTCGFSIGIRLCHICIGFGAMQFLGSPMAATKGLIADPEYPMDETEPRPA
PMAVYPHHEYLPIEERSNLNLSMDTAPLOKCYRISFQCRKYSIAGDPTLPS
TFDMAAHIMRSWAKLDVLPDYMLRTFVSVAATLETLLKRGFGVGLACAC
SSVESLINDSLKTRIVADARLRSVEDYLRSMVDYVKKRKLREBGRKLTITOMR
FEMEYTADEGNGKPYVAGPIDLRPTPOCVLLPQGVSGSDQSMVYCLCPPIAVHT
FTLLSLNDHK"
complement(join(6700..7357,7482..7884,8620..8839,
8920..9241,9758..10380))
/note="unknown protein: similar to ESTs gb|T04168.1,
gb|T21964.1, and gb|AW004462.1"
/codon_start=1
/evidence=not_experimental
/product="F21B7.3"
/protein_id="AAF86540.1"
/db_xref="GI:9280671"
/translation="MGHDSFCYLIVLRALRCGLIATLMOTCALOKKRRSMESDSS
SOAKAFDETKTGKVLASGIGKEIPAMPHRPDPLTSLKOTAPPSQOUTTPYDCLKG
SMDLISRSVYEKIGDAERGFQVYVHGISVYMERKMGIRFRHODEVKKRY
SRDTRVLYSNITDHTCKAKANRRPFLACVAMPDPKLODIPAVGGEIMMEYKOL
MTLGEPIFELLSALGLNPHLKDMGCAKSHIMGQYPPPOPDILGIGSKHDESF
ITILLQNIIGLOVYIDQCWVDVSPGALVYINGDLOLISNCKFISAHERVANS
SEPRISMPCFVSTPMKNPRIYGPRIKELISQNPARKRDILTFFSMTFSOTRIE
IPALFRATPATILASLKSPPPKHLITPVLDKGSVYEKIGEAEEKGLPHVNHGIP
VEVLERMIOGIRGHEOPEAKKRFYSRDRVLYSNINDLONSEASNRDITGCT
APBRPLEDIPAVGGEIMMEYKELISGERLELSALGLNHLKDMCASQVY
VGQHPRCRPODLITGINKRTDLSFLVYLQDNGVGLQVHEQWIDVPGALVYN
IGDFLOVKKIMLRLSCFISCMIMGLILLEYCVOLITNDPFISEHVIANGSEPT
SVAIVSTFMKAVSRVYGPIDKILSAENPAKRYDCITLFEISTFSSKTLDAPKLHHRK
I"
10833..11348
/note="unknown protein: similar to EST emb|230724.1"
/codon_start=1
/evidence=not_experimental
/product="F21B7.4"
/protein_id="AAF86509.1"
/db_xref="GI:9280640"
/translation="MRGFSVSSWMLMTSVIEIGVPLGISTTWLDVAVAVADAVMWFVS

CDS
ACLMANKDLRLNLRVPAEGLSESPSPFSLRFLSEISRIECCAFGRKFSDLRPMRSE
AKKCVVLEKMGWLDLQIGMRSVNMVFSVFFVAGPASPILRDLRODYSYISANDLVVVV
LQSSPCIRRY"
join(11965..12033,12173..12303,12448..12519,12596..12685,
12756..12915,12985..13041)
/note="similar to RHP1 dbj|BA37110.1"
/codon_start=1
/evidence=not_experimental
/product="F21B7.5"
/protein_id="AAF86510.1"
/db_xref="GI:9280641"
/translation="MNTIYVAQLOQOFDYIYSLYQOGLDNOSEELKLODECTPDF
VAEYSLFPDQSKLINTMSISLERPDVNDKQVDSGVHQLKSSSLICFVLSGAR
VANVLSIRECCDVNRGYSNNTNDNDIMRCLNCLQDVDEYKMLTKLODLF
NVSFLRCNLVLLEKQIILQAGTIPQVDIN"
16495..17682
/note="similar to protein kinase p1r|A57676, similar to
ESTs gb|A1993651.1, dbj|AV538995.1, dbj|AV522524.1, and
dbj|AV538995.1"
/codon_start=1
/evidence=not_experimental
/product="F21B7.6"
/protein_id="AAF86511.1"
/db_xref="GI:9280642"
/translation="MRFKYLWCMFLRFGFTTEALIDPVDFLALQAIKSLDLPLG
SKFESWDTSDPCGFAGVYONGKIVSLNIGDPRAGPGLSGIDPAIGKLSALTEL
SIVGRIMGALPATISQLDLRFLASINFTSGEPASIGVRGARTLDLSYNDLTG
ISPSIGSLPELSNLILCNHNLGTSLPELSTLRIDLRKNSLIGSLSPSLPQIG
LOPANOVITSDVLSYNSRSGISPLBSVNLILNSRFTGVEVPASIVETLANSQI
TLYLOHNLTGQISPAEIVSSSLQTNQCMVPLPOTPCPLAGDPKRPPTTQIE
WRG"
complement(join(18462..18515,18612..18742,18840..18975,
19044..19139,19214..19410,19497..19595,19668..19752,
19865..20103,20179..20253,20345..20649,20798..20902,
21088..21216,21356..21437,21513..21608,21780..21849,
21929..22025,22408..22502,22586..22659,22810..22888,
23041..23101,23206..23313,23411..23536))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F21B7.7"
/protein_id="AAF86539.1"
/db_xref="GI:9280670"
/translation="MFKKRRSLNFIICLPKKECARGOILGYKMETKAIIGPHNLS
PQNLVCGHTLVAEVNNSHOLILEGGSTTVANNSSLPEISLDGVTNVSFVLTFR
KMTRLNIGVSPRACHAALYGTLLIGIGISGSDGDMYMDMTNKKMKIETLY
GGETSPRYGHVMDIAORMLVIFSGNNGNLEIDDTWALDTRGPFSVDRINPSGNDP
GRMAYSGSREDGIFILCGIDHSGYTLGDTYGLKMSDVMVTPVPAVAFSPRQHTA
VFGSKLHVIGILINRRLIDGEAVVADTNQPEYSAGANRQNOYOLMRCHHAASF

Query Match 6.7%; Score 102.2; DB 7; Length 101176;
Best Local Similarity 45.4%; Pred. No. 8.4e-11;
Matches 620; Conservative 0; Mismatches 718; Indels 27; Gaps 6;

CDS
7 ACCGTCCTCGAATCTCTCCGAGTGGCGCGCTCCAGGCACGGTGGCTGACACTACTC 66
11
Db 37271 AACATCATCGACCTGTCACATCTCTCCCAAGGACCGCTTGATCCACCACTCTT 37330
11
Qy 67 CCGGTCACCTTCTTCGACATGAGCTGGCTGCATTTCACACCCATGCTTACACTCTCTG 126
11
Db 37331 CCTCTCACCTTCTTCGATGCCCTGGCTCTCTCCACACGCGCGAATCTCTCTTTTC 37390
11
Qy 127 TAGCAACTCCCGCTTCGCAACCGCGCTCTCGAACAACGGTCTGCCGAATCAACAA 186
11
Db 37391 TTCTCTTACAAAACAACTGAATCTTCTCCMAAGACTTGTACCACAACTTAAACAT 37450
11
Qy 187 TCCTTATCTCTAACCCCTCAACACTTCTCCCGCTTTCATGACATCTTAATCTACCCCTTA 246
11
Db 37451 TCCTCTTCATCATCTCTCCAACTTTTCTCCCTTACCGCGGTAACTGATTAATCCGCGCT 37510
11
Qy 247 TCGCCGAGAAATGCGGAGTTCCGGTATCAGAACGGTACTTCGGTTCTTTCACAGATT 306
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

Db 37511 GGTCCGACCCCTCCTATTTTACACTACACCGCTGGGAGAGACTCTCTTTTTCACCCTA 37570
 QY 307 ATGAGATCT-----GTGAGAGATCATCCGATTCGCTCATTAATTAATCTACTGCTT 357
 Db 37571 GCAGAGCTCTGTAACACGCTTGACCACTGAGTCTGATTCCTTAAGATATCACT 37630
 QY 358 GCCCTTAGCGACGATTAATGAGATCTCCAGCTGCCCGATAGTCGAGAAATCTGATCGG 417
 Db 37631 GTGTGGATGGCTCTTGGCCCAAGTATCCCTCTCCTCCTCCGGAAGAAATTCAG 37690
 QY 418 AAATGTGTTCAAGTTTATGACCGGAGTGCCTGTTCCGCTCCGCGGGGTGATC 477
 Db 37691 A---TGACACCAATATGGAATGCAAGTACCATCTTCCCTGAGCTGGAATCTATA 37747
 QY 478 GGAATTAACGACGACACACACCGTTAGCGATCTCATCTGTTGAGGTTTATGAAGAGT 537
 Db 37748 GGCACACTCAGCTACATGTTGTAGAGAGATGAGTCCCTTCACTTATTCATGAAGTAT 37807
 QY 538 TGGGCTTCCTCATCAATAATTCGAGAGATGATGAAATCTTGGACGGAAGGTAATGT 597
 Db 37808 TGGATGTCATGACCAATTCAGCGGCAAGATCCGGCCACGGTCTTCTACCTCTCTG 37867
 QY 598 TTGCGGTTTTGACACCGATGCTGATGATTAATCCGCTAAATTTGACACATATTTATG 657
 Db 37868 CCTATTACAGCTGACGAAACATATACAGACCCAGAGAGTACGCGACGACATTTG 37927
 QY 658 AACACCGCGAGAAAGCTCCGTTGGAATCCAGCATCCATCTTTACCGAGATCCGAT 717
 Db 37928 GACGCGTTTTGAGCCAAACTCGCAAAACAGTTCACATGTTACGCTGAGAACATG 37987
 QY 718 C---GAGCTACCTACTTTTCCACCAATCTGAATTAAGAATTTGAAGGTTGATTCAG 774
 Db 37988 GTGAGAGCCCTTTTACATTTAGACCCGGAACAGATACATCTCAAGAGTTGGGTACA 38047
 QY 775 AGAAAGCCCCCAATGATTAATCTCTCTGCTGCGCATTCGAGGTTATATCTG 834
 Db 38048 GACAGCTGAGATCAATCTCTGTTTCTACCTTCTGCTGATCTGATCTTCAATTTG 38107
 QY 835 ACCGCGATCCCAATTCGTCGAGATTTACA---AAGACGTGATGACGACAAACGCGCT 891
 Db 38108 GTAAAGCTGATTAAGACACTTTTCAAGACAGTGAACAGAGGCAAGGCAAGTCTT 38167
 QY 892 TTTCTTTTAATTCGATCGATTTAAGCGCCGCTTTGA-----TCCGCGGCTCCGGG 945
 Db 38168 CACTTGTATGATCAACGAGTTCAGGAATCTCTCAAGTACACACAAACCAATACCA 38227
 QY 946 AACTACTTGGAAATGCTCATGTTGGATGGCGGAATCTCGCGGGGATTTGCTC 1005
 Db 38228 ACATTAATTTGGCAACTGATAGCTCTGATGATCTCAAGAAACAGATTTGCTA 38287
 QY 1006 GGAATGCAAGGGTGTTCGGGAGCTGAGCGATCGCGCGGAATATGAGAAGAGAGC 1065
 Db 38288 GGAAGAAAATGCGTCAATGGCTGCTTCAGATGCAATCAGCAGAAATCAAGATATG 38347
 QY 1066 AGGCAAGAAAGATTTAGAACTGTGAGAACTGGCCGCTGATGATTCGCGAAGCTTG 1125
 Db 38348 TCAAGGATCTGTTGAAGACAGACCAAGATGGGACAAAGATACCTAATATGGTCTAG 38407
 QY 1126 CAATACTGTAATTCGCGTGGGGGATCGAGAGCTTATCTTTACGCGCGGATTT 1185
 Db 38408 TCCATTAATCCAACTCAATTTGAGGAGCTCCAAATTTGGAGCTGATGATGATGATG 38467
 QY 1186 GATGAGGATGAGGCGGTGAAGCAAGATATCTGATGATGAGAGAGATTTACAGT 1245
 Db 38468 GATTTAGGAAAGCGCTGCAAAATATGAGATATGACATAGAAACAG---GTGGTTCTATC 38524
 QY 1246 TCGTTGTGAACCGAGGAGTCTGCCGAGAGATTTGAGGTTGATTTGTTGCCAAAG 1305
 Db 38525 GCATTTCAAGAGTCCAGAGCGGACGACGAATGATGATGATGATGATGATGATGATG 38584
 QY 1306 GAGGAATTCGAACCTTTTGTATATTTTGGCGAGGAGATTAAG 1350
 Db 38585 AAGAAATGATGTGTTTGTACTCACTTCTACAAAGGAGGATCAAG 38629

RESULT 10
 AC003027
 LOCUS
 DEFINITION
 Arabidopsis thaliana chromosome I BAC F21M11 genomic sequence.
 AC003027
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 119914)
 REFERENCE
 AUTHORS
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremetska,I., Kim,C., Lenz,C., Li,J., Liu,S., Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaya,V.S., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 Unpublished
 2 (bases 1 to 119914)
 JOURNAL
 REFERENCE
 AUTHORS
 Federspiel,N.A., Palm,C.J., Conway,A.B., Kutz,D.B., Conway,A.R., Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Ojima,O., Osborne,B.I., Shinn,P., Sun,H., Toriumi,M., Vysotskaya,V., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 Direct Submission
 Submitted (22-OCT-1997) Biochemistry, Stanford University/DNA Sequencing and Technology Center, 855 California Avenue, Palo Alto, CA 94304, USA
 3 (bases 1 to 119914)
 REFERENCE
 AUTHORS
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremetska,I., Kim,C., Lenz,C., Li,J., Liu,S., Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaya,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 Direct Submission
 Submitted (30-DEC-1998) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
 TITLE
 JOURNAL
 TITLE
 JOURNAL
 COMMENT
 On Dec 30, 1998 this sequence version replaced g1:7234094. Bases 1-9262 of clone F21M11 overlap with bases 68998-78259 of 'RAMU' BAC clone F20D22 (AC002411) and bases 119525-119914 of clone F21M11 overlap with bases 1-389 of 'RAMU' BAC clone F21B7 (AC002560).
 e-mail for correspondence: arabseq@stanford.edu
 Genes with similarity to proteins in the databases are described as 'putative', 'like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'.
 The software programs used to predict genes include: Grell (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://genome.stanford.edu/~chris/GENSCAN.html), Fexa (V.Solov'yev & A.Salamov, Sanger Centre, http://genome.sanger.ac.uk/), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).
 Location/Qualifiers
 1..119914
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"

FEATURES
 source

```

/db_xref="taxon:3702"
/Chromosome="1"
/clone="F21M1.1"
1. 7000
/note="Proteins in this region are annotated in the F21M1.1. 9262
entry, AC002411."
misc_feature
/note="Overlap with bases 68998-78259 of 'IGF' BAC clone F21M1.1"
7684..10913
/gene="F21M1.1"
join(7684..7666,7965..8027,8122..8160,8250..8381,8458..8728,8875..9893,10093..10256,10344..10410,10501..10551,10685..10781,10885..10913)
/gene="F21M1.1"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AADI0665.1"
/db_xref="GI:4204284"
/translation="MAPPLRIKKACDANKLIGSEPTKRAFLRLKLTITENNDITE
DAKVLDAIFDEADOSTEKNNKKEEKKKEKSSVATSGRRKAEPLVODEE
DDMDDEEPLKRLRSRRGRASSSSSSNNNDLTOPEEDEDGVELPLKRY
VRRNERGLAMTVYNNASPSRSMRREVPFVLLPAPMETKYSEASALVIND
EPNIDHKPVI SDTGNCSAPMLMKKSNH VQEWMEKDLINDTADVPSSAIGES
SEHKVAASVELASTSGEATICSFAFATGETTNLHLPSEDLRRAMEKCKSKI
VHPEFVSGFPMKDKSCYIDLAKNSTQLLETIVCDMSKADSGAVGIMLVYVP
BCETISGDGKRAISNKKDITAGEENVEIIPWNEIENKVSFRFMPHSPVODAPVIES
LSSPDESGCSTSCIEDCLASEMSCNCAIGVNGFAVTDGLKKEPLRARSADQ
KVOYLRFCEECPLERARKVELLEPCGHLKRGAIKECWCCTKCKGRNVORGDH
KLOVFFPNKGWGLRTLEKLPKGAFCYEYIGELTTPELYXSFEDKPLVILDAH
WSEERLEGDKALCIDMGFYGNISRFNLHRCIDANILEPVOETPDHNYHLAFETT
RDIAMEELAMDYIDFNDNDSLMKPDCYCGSRFCRKNKRSKTKMOINKA"
join(13033..13488,13717..13926,14139..14514)
/gene="F21M1.2"
13033..14514
/gene="F21M1.2"
join(13061..13488,13717..13926,14139..14316)
/note="Similar to acid phosphatase; location of ESTs 110C277, gb|T42036, and 110C2Xp, gb|A1100245"
/codon_start=1
/protein_id="AADI0666.1"
/db_xref="GI:4204285"
/translation="MDRTMFLSLTASLIVGVVSAGDNILNOLRGSSONGIYS
KGIKIDLKGYCSMRINVEVNIIRKFDVPOCVSHIDYMTSSQIKODVARTDEVY
LHFGSMCKSKSGCDMAVIDIDITLSTPYKKNCFEGEKLNSTKEDVARTDEVY
APAVPHMKLYHDIERGKIIFLISREKYLRSATVDNLIOAGYGSNMLRGLEDO
OKEVQYSEKSKWMLSGYRWGVMGQMSFAGCPLPRTFFLPMNSIYYVA"
complement(join(14004..14312,14565..14641,14787..15831,15906..16006,16327..16362))
/gene="F21M1.3"
complement(14004..16362)
/gene="F21M1.3"
complement(join(14567..14641,14787..15831,15906..16006,16327..16362))
/gene="F21M1.3"
/note="Unknown protein; location of ESTs 40C37T, gb|AA728590 and 40C377, gb|T04573"
/codon_start=1
/protein_id="AADI0667.1"
/db_xref="GI:4204286"
/translation="MLDQISGNTHEKLSYEAEPHILNDRYHVEIIPKASVIDITE
ICDAEERKCSPTISRKRYTDFSKVTEHYVSESEVSEKNEVESEKSLSKS
TDDOIILEVANSNGSYDENHRYKNCRESDDIIEEDFSDSDLEDESEKSLSKS
DSLHNPTEKYTODIGDKTEIEIDSKLRNENYVRDNDHGGGVANPENLTKOMSAK
LAVDASLSTMLSTSESGSCNLSASMTTLPEKLSKSCYSKRLRNHNDRLVLAITL
EDIKOPSATSTPKSPKSPDEPPIIGVGYWGRNSKALIDGSSASPRKGIPTNSKTI
REKSYVMHSTPEARLEKALNNIDK"
complement(16835..20238)
/gene="F21M1.4"
complement(join(16835..17185,17274..17392,17491..17656,17795..17885,17982..18079,18173..18361,18504..18604,18705..19049,19134..19349,19439..19791,19862..19967,20188..20238))
/gene="F21M1.4"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AADI0669.1"
/db_xref="GI:4204288"
/translation="MGSSSPPEARARAVPSMILIFELICTVHYVYTNBRKLDVLSA
NLNIPKRYTRRDLRFMEVSIISYKSLAAVSHLPOLOIPNCCNKNAKLNKNSPKH
GESDESDMDKDKSKSGSDSSRSGSPLPSESDPRPHODTEKQLSDHLLYE
FSESEYAAHNTPESEYEOAKKVVRODITTAESOPSNARRIGDSSTOESSNPMDOP
TLRLMESLRSDPDYVYAKOHQWVPSFWYTFYFVLCLSNLRCAVIFCVQOPLDP
SHTEODSKRRDDITADAMENHLKVPKRENNIMQMSAIDIDNGCSNADSDOLSEKS
KALEOTSSNITIGCPOASARVEATGEMHLISRGHPVDGDIIFRSNJIHVISACIEVA
POVYVGVDPVKKLKAELAKMKTICCTCSLKGALGCPVKSCRSYVPCARISRCR
MVEEDFLICPAHSSVYKTFNEKSGHVRARPELPKINPAELCSLEOTPAFLKLVLCG
JLNAAMKASLKSASQPVDEEPELIDIDGCGDDPKTARLARAEFKRLKVLGMLNGM
GDFYTGVEEDLQNLVAVGTLINTEDELAGESSNNDORSSITVYTNIDPPHGAL
GEVYTIWQRANDAEALASQTSRLVGHVTLWLESIAQKLPVYG"
/gene="F21M1.5"
join(20752..20994,21362..21497,21596..21740,21825..21994,22102..22178,22542..22758,22920..23064,23343..23442,23599..23693,24021..24072,24227..24298,24441..24647)
/gene="F21M1.5"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AADI0668.1"
/db_xref="GI:4204287"
/translation="MGANSKSVASFVYVAVFLLGCGTFAVEDETEFHGYSKLSGI
IIPGFASSTOGLRANSIILDCPYTPIDPLPDLVWLDTTKLLSAVNCFKCVLDPVNOTD
HPECKSPDSGLSAITLEDPGYITGLSTVYMEWLMKVEFGIENANAIYAVPYMRLS
KPLDQNHATFAFGAPLGSVEALIKSTLGSVFGVGLSGVSTGRNLIENSPASLHMPF
SKCKGDNFTWFHSGGAARKDKRKRVYHCEEEYOSKSPRNIIINTIPSTASBELA
DGLFLFAIEDYDPDSRMLHQLKVKVPFVIRNIRSLAGFLIYHDPVFNPLTPW
ERPIKRVICITGAHLKTEGVYFAPASCPYDNNIITDITIEGSLVSGSVVVDG
NAGPIGTDETVPHSLSMCKNMLGPRVNTKAPOLILKIKQEPHDSDVHVELND
HEGSDIILNMTKAPRVKITYTFEDESIPCKRTYVWELDKSGY"
/gene="F21M1.6"
27777..28734
/gene="F21M1.6"
28007..28465
/note="Unknown protein; location of ESTs 203I247T, gb|H76794 and 203I24XP, gb|AA605510"
/codon_start=1
/protein_id="AADI0670.1"
/db_xref="GI:4204289"
/translation="MNTKTYMLPPRRVLFADKREKRPASISVTNDNPELAFPSPP
KLVPVPVNDISKSSSTAABEIGSNOLMLAGYLSHETLUGTIFGSGWQANRAOAESS
KIKPSHYEPAECEPEPKRRRYREAVANLRSBDGQDLPEIVNPAOLARFLK"
complement(129264..32033)
/gene="F21M1.7"
complement(join(29264..31015,31312..31414,31484..31587,31782..32033))

```

```

Query Match      5.88; Score 87.8; DB 7; Length 119914;
Beat Local Similarity 44.98; Pred. No. 7.3e-08;
Matches 617; Conservative 0; Mismatches 717; Indels 39; Gaps 6;

OY 11 TCCTCGAATCCTCCCGATGCGCGGCTCCAGCAGCGTGGCTGAGCAGTACTCCCGC 70
    || ||||| ||| | | | | | ||||| || | | | | | | | | | | | | | |
Db 45957 TCATCGAGACCTGTCACATCTCTCCCAAGGACCGCTCCATCAACACACTCTCC 1
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 71 TCACCTTCCTCGACATAGCTGCCTGCATTCCACACCAATGCTTCAAGCTCTCTAC 130
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 46017 TCACCTTCCTCGAAGCCCTCGCTCTCTCCACATGCGCAGATCTCTCTCTCT 1
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 131 AACTCCCTTCCTCAACCGCCTTCTCGAAACGCTCTTCTGAAACCTCAACCAATCT 190
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 46077 CTTACCAAAAACCTGAAGTCTTTCCCTCCAGACTTTGTACCCAACTTCAAAACATTCCC 46136
OY 191 TATCTGTAACCTTCAACAACTTCTCCCTTTTCATGCAATTAATACCTCTATCGC 250
Db 46137 TCTGCATCACTCTCCCAAACTTTCTCCCTTACGCGGCTAAAGATTAATCCCGCTGTC 46196
OY 251 CGAGAAATATGCGGAGTCCGATATCAGAACGCTGCTGGTTTCTTTCACGATTATYGC 310
Db 46197 CCGACCCCTCCATATTTACATACAGAGATGGCCAAAGACTCTTGTATTCACCGCTACAG 46256
OY 311 AGCTCT-----GTGCGAGATCATCCGCAATCCGCTCAATAATACTACTGTTGCC 361
Db 46257 AGCTCTACTGAACCGACTTCCAGCAACTGAAGTCTGATTTCCCTAAAGATATCAGTGT 46316
OY 362 CTAGGACGATTTATGAAGATCTCCAGCTGCCCGCATATGTCAGAACTGATCGGAAT 421
Db 46317 TGCATGGCGCTTGTCCCAAGTTACCTCCCTCAGCTCTCTCGGAAGGAATTCAG --- 46373
OY 422 TGTTCAGATTATAGCGCTGCAAGTACTGTGTTCGCCGCTCCGCGGCTGCATCGGAA 481
Db 46374 TGGGACCAATTAATGGCTATGCAAGTCAACCATCTTCCCTGGAGCTGCAATCTGATAGCA 46433
OY 482 TACGACGCAACACACGCTTAGGATGCTCCATCTGTTTAGGGTTTATGAAGTGGG 541
Db 46434 ACTGACCTACACATGTTTGAAGAGATGAGTCACTTTCAGTCAATTTTCAATGATTGGA 46493
OY 542 CTTCATCACTAATTAATTCGAGAGATGATGATTTCTTGAAGGAAAGTGAATGTTTC 601
Db 46494 TGTCACTTGAACCAATTCAGGCGGTAAAGATCCCGCACGGTCTTCTTACCTCTCTGCTTA 46553
OY 602 CGGTTTTCGACCGATGCTCGTGAATTAATCCGCTAATTTGAACACATATTTATGAACA 661
Db 46554 TTTCACAGCTGCAGAAACATGATCAAGACCCAGCGAGTAGCGGACGACATTTGGAGC 46613
OY 662 AGCGGAGAAACCTCCGCTTGGATTCGACAGATCTCATCTTACGAGGATCGCATTC --- 718
Db 46614 GGTTTTGGACCAAACTGCAAAACGATTCACATGTTTACGCTCGAGAACATGCTCA 46673
OY 719 GAGCTACTACTTTCACCCATCTGAATTAAGAAATGAAGGGTTTATTCAGAGAA 778
Db 46674 GAGCCACTTTTACATTTAGCGGAGACAGATTAATTAATTAAGATTTGGCTTACAGAGC 46733
OY 779 AAGCCCAATGATGTTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 838
Db 46734 AGCTGAGATCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 46793
OY 839 GCATGCGCAATGCGTGGAGATT-----CAAGAGCTGATGAGACACA 883
Db 46794 GCTTGATCAAGACACTTGTACAAAGACAGTGAACAAAGGCCAAAGAGAGACAGAGATG 46853
OY 884 AAGCGCTTCTTCTTAAATTCGATGATTTAAGCGCGCTTTGGAT-----CCGCGG 937
Db 46854 AAGTATTTCACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 46913
OY 938 CTCGCGGGAATCTCTTGGAACTGTCTATCTGTTGGATGCGGAAGATCTCTCGCGGG 997
Db 46914 TACCAAAACATCTTGTGGCACTGATGCTCTGCTGATGATGATGATGATGATGATGAT 46973
OY 998 ATTGTCGAGATGAAGGGGTCTTTCGCGCACTGAGGAGATCGCGCGGAAATAGAGA 1057
Db 46974 ATTGTCGAGAGAAAATGCTTTTGGCGCTTCAAGTGAATCAACAGCAAGATCAAG 47033
OY 1058 AGAGAGAGAGAGAGAGATTTCTAGAACTGTGAGAACTGCGCGTCTGAGATTCTGCG 1117
Db 47034 ATATGTTATCAACGATCTCTTAAGACAGCACCAGAGATGGGAAAGAGATCTGATTAAT 47093
OY 1118 AAGCTTTCGAAAATCTTATTTCTCGGTGCGGAGATCGAGCAGGCTGATCTTTACGGCG 1177
Db 47094 GGGTCACTGCCATTTATCCAACTCAATTCAGAGAGCTCGGAATTTGGAGCTGATGACA 47153
OY 1178 CGGATTTTGGATGGGTAAGGCGGTGAAGCAAGATGATGATGATGATGATGATGATGAT 1237
Db 47154 TGGATTTTGGATTAAGGAAGCCCTGCAAAATGAGATCTGTACATGAGAAACAG ---GTG 47210

```

```

OY 1238 TTACGATGCTGTTGTAAACCGAGGATGCTCCGAGGATTTGGAGTTGATTTGCTT 1297
Db 47211 GTTCTATCGCATTTCTCAGAGCCAGACAGCAGCAATGAGGTTGATTTGATATAC 47270
OY 1298 TGGCAAGAGAGAAATTCAGAGCTTTGATGATTTATTTTTCGCGAGGAAATAG 1350
Db 47271 TAGAGAAAGAAATGAGATGTTGATGATTTGATGATTTTCAACAAAGGGGATCAAG 47323

RESULT 11
AB028618
LOCUS AB028618 85690 bp DNA PLN 20-NOV-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone:MOD1,
complete sequence.
ACCESSION AB028618
VERSION AB028618.1 GI:5041971
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:11b:Mitsui P1
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids. II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Nakamura,Y.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. II
JOURNAL unpublished (1999)
REFERENCE 2 (bases 1 to 85690)
AUTHORS Nakamura,Y.
TITLE Direct Submision
JOURNAL Submitted (09-JUN-1999) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2, 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail:ynakam@kazusa.or.jp, Tel:+81-438-52-3935,
Fax:+81-438-52-3934)

FEATURES
source
1. 85690
location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MOD1"
/clone_11b="Mitsui P1"
BASE COUNT 27956 a 15199 c 15751 g 26784 t
ORIGIN
```

```

Query Match 5.1%; Score 77.8; DB 7; Length 85690;
Best Local Similarity 48.5%; Pred. No. 7.7e-06;
Matches 276; Conservative 0; Mismatches 287; Indels 6; Gaps 2;
```

```

OY 910 GATTTAAGGCCGCGTTTGAATCCGCGGCTCCGAGGAATCTTGGAAATCTGTAATCG 969
Db 16051 GATTTCAAGAACCGGCTAGATTCGGAAGCTCCCGGGGACTTTCGGAATCTGATGAT 16110
OY 970 TTTGCGATGGCGAAAGATCCGCGGCGGATTTGTGCGAATGAAGGCGTGTTCGGGCA 1029
Db 16111 CCGGTGGCTGTTATTAACCGTAAGCGCGGAGGTTATGGAAGAGAGGATTTGTACG 16170
OY 1030 GCTGAGCGATCCGCGGGAATAGAGAAAGAGACAGCAGCAAGAGATTTAGAAACT 1089
Db 16171 GCGGCTGAGATTTATTAAGCAATTTGGTCAAAAGGCTGAGTTCAGAAAGAGATGAGACATT 16230
OY 1090 GTGGAGAACTGGCCGCTGTGATTCGCG ---AAGCTTCGAAAATCTGTTATTTCTCGCG 1146
Db 16231 GCGGACACTTTCGTGGAAGGATTTAGTTTCAAGATTTGAGTGTACACAGATTGGAGCAT 16290
OY 1147 GCGGAGTCGAGAGCGCTGATCTTTACGCGCGGATTTGGATGGGTAAGGGGCTGAAG 1206
Db 16291 GCGGCGTGCAGCCGCTTGGAGTATACGAGCGGATTTCCGGGTGGGAAAGCCGTTAAA 16350

```

Qy	1207	CAGAGCATCTATGTCGATGATGAGAGAAATTACGATGTCGTTGCTAACCAGGAGAT	1266
Db	16351	GTTGATATATGTCATCATGATGACCAAGAGA---AGCATGCGATGGCTACAGACGTTGAT	16407
Qy	1267	GCTCCGAGAGATATGAGATGATGCTGTTCCCAAGAGCAATTCGACACTTTTGAT	1326
Db	16408	GAGTCAGAGTGGCGCTTACAGATTCGATGTTGAAAAACATGAATGATTCGGTCGTG	16467
Qy	1327	GATTATTTTGGGAGGAAATAAGGGTGAATTAATCATTTATCATGATATGAGTGTG	1386
Db	16468	TCTTTTTCACAAATGGTTTACATAGCTAAACTGTTGTTTGTAGATTTTAATTAATAG	16527
Qy	1387	GATGAATACCTCTGTTTCATCTCATTTGTTTAAACAATTAATTTTCCATGAACCTTT	1446
Db	16528	GTAATGCTTGTTGATTTATCTGTTTTCCTCACAAGAGATTTGGGCATTAATCTCC	16587
Qy	1447	TTCATCATTAATAAAAAAAAAAAAAAAAAA	1475
Db	16588	ACTAATTAATTAAGTTTAATTAAGATATA	16616
RESULT	12		
AC069470	136047	bp	DNA
LOCUS	Arabidopsis thaliana chromosome I	clone IGF-F7F7,	28-JUL-2000
DEFINITION	PROGESS ***	4 unordered pieces.	*** SEQUENCING IN
AC069470	GI:9558603		
VERSION	AC069470.8		
KEYWORDS	HTG; HTGS; PHASE1.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 136047)		
AUTHORS	Lin, X., Fujii, S., Town, C.D., Benito, M., Ceasary, T.H., Rouning, C.M., Koo, H., Kull, C.Y., Utterback, T., Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.		
TITLE	Arabidopsis thaliana 'IGF' BAC 'F7F7' genomic sequence near marker 'mi72'		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 136047)		
AUTHORS	Town, C.D. and Kaul, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@ligr.org		
COMMENT	On Jul 28 2000 this sequence version replaced g1:9440543.		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 4 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	1 100163: contig of 100163 bp in length		
	* 100164 100213: gap of unknown length		
	* 100214 112623: contig of 12410 bp in length		
	* 112624 112673: gap of unknown length		
	* 112674 116636: contig of 3963 bp in length		
	* 116637 116686: gap of unknown length		
	* 116687 136047: contig of 19361 bp in length.		
FEATURES	Location/Qualifiers		
SOURCE	1. 136047		
	/organism="Arabidopsis thaliana"		
	/db_xref="taxon:3702"		
	/chromosome="I"		
	/clone="IGF-F7F7"		
BASE COUNT	46028 a 22872 c 22213 g 44783 t 151 others		
ORIGIN			
Query Match	5.1%; Score 77.8; DB 59; Length 136047;		

Best Local Similarity	48.5%;	Pred. No. 8e-06;	
Matches 276;	Conservative 0;	Mismatches 287;	Indels 6; Gaps 2;
Qy	910	GATTTAAGCCCGCGTTTGATCCGCGCGCTCCCGGAGACTGCTGGAAAGTGTCTATCG	969
Db	90952	GATTTCAGAGACCGGTTAGATCCGAAGTCCCGGGGACTTACTTCCGAAACGTGTATGAT	91011
Qy	970	TTTCCGATGCGAAGATTCCTCCGGGGGATTTGGTCGAGATGAAGGGGTGTTGGGCA	1029
Db	91012	CCGCTGGCGCTTATTAACCGTAAGCGGGCGAGTTTATGGAAGAGAGGATTTGTGACG	91071
Qy	1030	GCTGAGCCGATCGCGCGGAAATAGAGAGAGAGAGAGAGACGACAGATCTTCNAAACT	1089
Db	91072	GCGCGTGAATATTAATGACATTTGGTGAAGGTTGAGTTCAAGAAATAGAGACCAT	91131
Qy	1090	GTTGAGAACTGGCCGCTGAGATTTGGCG---AAGCTTGCAAACTGTTATTTCTCGGTG	1146
Db	91132	GCGGACACTTTGTTGGAAGATTTAGTTTTCAGATTTGAGATGACACAGTTTGGAGCAT	91191
Qy	1147	GCGGATCGAGACGCTTGATCTTTACGCGCGGATTTTGGATGGGTAAAGCGGTGAAG	1206
Db	91192	GCGGATCGACCCGTTGGAGTATACGAGCGGATTTCCGGTGGGAAACCCGTTAAA	91251
Qy	1207	CAGAGATTAATCTGCTGATGATGAGAGAACTTACATGCTGTTGTGTAACCGAGAT	1266
Db	91252	GTTGATTTGCTCTCATTTACCAAGAGA---AGCGATCGCAATGCTGAGACGCTGAT	91308
Qy	1267	GCTGCCGAGAGATGAGAGTTGATTTGCTTTGGCCAAAGAGCAATGCAAGCTTTTGAT	1326
Db	91309	GAGTCAGGTGGCGTTGACATTTGGAATGTTGAAAAGACTGAATGATGCTGCTGTC	91368
Qy	1327	GATTATTTTGGCGGAGGAATAAGGTTGATTAATCAATTAATCATGATTAATGAAGTTC	1386
Db	91369	TCTTTTTCACCAATGCTTTACATGATTAACCTGTTTGTGAAGATTTTAATTAATAG	91428
Qy	1387	GATGAATCTCTGTTTCATCTCATTTGTTTAAACAATTAATTTTCCATGAACTTTT	1446
Db	91429	GTAATGCTTGTTGATTTATCTGTTTTCACACAAGAGATTTGGCATTAATTTCTCC	91488
Qy	1447	TTCAGTCAATAAAAAAAAAAAAAAAAAA	1475
Db	91489	ACTAATTAATTAAGTTTAATTAAGATATA	91517
RESULT	13		
AP000606/c	82360	bp	DNA
LOCUS	Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MTO24,		
DEFINITION	complete sequence.		
ACCESSION	AP000606.1		
VERSION	GI:6045161		
KEYWORDS	HTG.		
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mtsui pl clone:MTO24.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (sates)		
AUTHORS	Nakamura, Y.		
TITLE	Structural Analysis of Arabidopsis thaliana Chromosome 3. II		
JOURNAL	Unpublished (1998)		
REFERENCE	2 (bases 1 to 82360)		
AUTHORS	Nakamura, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-OCT-1999) to the DDBJ/EMBL/Genbank databases. Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakam@kazusa.or.jp, URL: http://www.kazusa.or.jp/gene-s2/, Tel: 81-438-52-3935, Fax: 81-438-52-3934)		
FEATURES	Location/Qualifiers		
SOURCE	1. 82360		

```

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="W7024"
BASE COUNT      27083 a 14790 c 14375 g 26112 t
ORIGIN

Query Match      4.6%; Score 69.6; DB 8; Length 82360;
Best Local Similarity 52.0%; Pred. No. 0.00036;
Matches 181; Conservative 0; Mismatches 164; Indels 3; Gaps 1;

QY 1139 TCTCGGTGCGGGATGACAGCTTATCTTATGACCGCGGATTTTGGATGGGTAAGG 1198
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15417 TATCGGTTCTGGGTGACCGGTTAGGATTTACGGTCTGATTTTGGGTGGGAAAC 15358
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1199 CGGTGAAGCAAGATACATCTGATTTGATGAGAGAGATTTAGATGCTGTGTAAC 1258
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15357 CGGTGAAGTAGATGATTTGATGATTTGAT--AAGATGCTTTCGTTTCATTTGCGAGA 15301
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1259 CGAGGATCTCGCCGAGATTTGAGATTTGCTTTGCCAAGAGAAATTCAG 1318
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15300 GTGGAGATGGAAGTGGTGTGAGTGTGTTGTTGAAGAAAGATGATGTGAAC 15241
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1319 CTTTGTATGATTTATTTTGGGAGCAATTAAGGTTGATTAATCATTTATCATGATTA 1378
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15240 GTTTTGTTCATTTATTTCTCATCGGTTGGAATAAATGAGACGTTTGGTTCTAGTT 15181
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1379 TGAAGTTGATGAATATCTCTGTTTCATCTCTATTTTAAACAATATTTTTCATT 1438
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15180 TATGTGTTGTTTATTTTCAATCATCTTTAATTCATTACTATTAGCAAGATGTGA 15121
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1439 GAACCTTTTGTGACTCATTAATAAAAAAAAAAAAAAAAAAATGAAA 1486
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15120 GCTGAAGCTGTGACATGAGAGAGATGAATTCGAGAGATGACA 15073
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
AC074226/c 166884 bp DNA HTG 20-JUL-2000
LOCUS
DEFINITION
IN PROGRESS *** 8 unordered pieces.
AC074226
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 166884)
Lin.X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Rouning,C.M.,
Koo,H., Fujii,C.Y., Utterback,T., Barnstead,M.E., Bowman,C.L.,
White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana 'IGF' BAC 'F27B9' genomic sequence near marker
'7'
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished
2 (bases 1 to 166884)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (20-JUL-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA. cdtoner@igf.org
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 10514: contig of 10514 bp in length
```

```

* 10515 10564: gap of unknown length
* 10565 31604: contig of 21040 bp in length
* 31605 31654: gap of unknown length
* 31655 37630: contig of 5976 bp in length
* 37631 37680: gap of unknown length
* 37681 44842: contig of 7162 bp in length
* 44843 44892: gap of unknown length
* 44893 69074: contig of 24182 bp in length
* 69075 69124: gap of unknown length
* 69125 135012: contig of 65888 bp in length
* 135013 135062: gap of unknown length
* 135063 159445: contig of 24383 bp in length
* 159446 159495: gap of unknown length
* 159496 166884: contig of 7389 bp in length.
FEATURES
source
1..166884
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="IGF-F27B9"
BASE COUNT      55357 a 28890 c 29784 g 52486 t 367 others
ORIGIN

Query Match      4.6%; Score 69.6; DB 60; Length 166884;
Best Local Similarity 52.0%; Pred. No. 0.00038;
Matches 181; Conservative 0; Mismatches 164; Indels 3; Gaps 1;

QY 1139 TCTCGGTGCGGGATGACAGCTTATCTTATGACCGCGGATTTTGGATGGGTAAGG 1198
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 111057 TATCGGTTCTGGGTGACCGGTTAGGATTTACGGTCTGATTTTGGGTGGGAAAC 110998
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1199 CGGTGAAGCAAGATACATCTGATTTGATGAGAGAGATTTAGATGCTGTGTAAC 1258
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110997 CGGTGAAGTAGATGATTTGATGATTTGAT--AAGATGCTTTCGTTTCATTTGCGAGA 110941
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1259 CGAGGATCTCGCCGAGATTTGAGATTTGATTTGCTTTGCCAAGAGAAATTCAG 1318
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110940 GTGGAGATGGAAGTGGTGTGAGTGTGTTGAGTGTGTTTGAAGAAAGATGTGGAAC 110881
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1319 CTTTGTATGATTTATTTTGGGAGCAATTAAGGTTGATTAATCATTTATCATGATTA 1378
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110880 GTTTTGTTCATTTATTTCTCATCGGTTGGAATAAATGAGACGTTTGGTTCTAGTT 110821
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1379 TGAAGTTGATGAATATCTCTGTTTCATCTCTATTTTAAACAATATTTTTCATT 1438
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110820 TTAGTGTGTTTATTTTCAATCATCTTTAATTCATTACTATTAGCAAGATGTGA 110761
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1439 GAACCTTTTGTGACTCATTAATAAAAAAAAAAAAAAAAAAATGAAA 1486
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110760 GCTGAAGCTGTGAGACATGAGAGAGATGAATTCGAGAGATGACA 110713
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
PFMALP4/c 224448 bp DNA HTG 11-AUG-1999
LOCUS
DEFINITION
IN PROGRESS *** 1n unordered pieces.
AL035477
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 224448)
Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
and Barrell,B.
Direct Submission
Submitted (24-FEB-1999) P. falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
On Aug 12, 1999 this sequence version replaced gi:5531400.
For more information about this sequence or the Malaria Project,
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 03:14:30 ; Search time 151.87 Seconds
(without alignments)
3764.790 Million cell updates/sec

Title: US-08-894-356C-6
Perfect score: 1522
Sequence: 1 NTGACGACCCCTCCGATC.....TTTTTTTTTTTTTTTTTTTT 1522

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.GeneSeq_36:*

- 1: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1505	98.9	1518	17	T37313 Aromatic acyl tran
2	765.4	50.3	1479	17	T37311 Aromatic acyl tran
3	196.2	12.9	1703	17	T37308 Aromatic acyl tran
4	174.2	11.4	1508	17	T37312 Aromatic acyl tran
5	153.8	10.1	1622	17	T37309 Aromatic acyl tran
6	62.4	4.1	4999	20	Z23902 Human LOBO homolog
7	55.4	3.6	4999	20	Z23902 Human LOBO homolog
8	55	3.6	2418	13	O27886 P.falciparum GBP13
9	54	3.5	2662	10	N90703 Rhodospirillum rubrum
10	52.6	3.5	1875	18	T48669 Human EDG-2 receptor
11	51.4	3.4	3138	12	O11712 Shuttle vector PMO
12	51	3.4	8365	20	Z20056 Plasmodium falciparum

C	13	50.8	3.3	1052	10	N90224
C	14	50	3.3	1518	17	T37313
C	15	50	3.3	5173	18	T89783
C	16	49.8	3.3	1132	20	X80658
C	17	49.8	3.3	2229	21	A23425
C	18	49.6	3.3	3138	12	O11712
C	19	49	3.2	4590	7	N60472
C	20	48.8	3.2	2418	13	O27886
C	21	48.4	3.2	1493	11	O04492
C	22	48.4	3.2	1493	20	X33994
C	23	48.2	3.2	4590	7	N60472
C	24	48.2	3.2	5852	12	O11710
C	25	47.6	3.1	198	14	O42784
C	26	47.4	3.1	1521	20	V73000
C	27	47	3.1	2662	10	N90703
C	28	47	3.1	2780	21	A26422
C	29	46.8	3.1	630	18	T79906
C	30	46.8	3.1	860	13	O22007
C	31	46.8	3.1	1864	8	N71405
C	32	46.8	3.1	4818	18	T62177
C	33	46.8	3.1	4818	18	T79909
C	34	46.8	3.1	5852	12	O11710
C	35	46.6	3.1	259	18	V00423
C	36	46.6	3.1	886	20	X51754
C	37	46.6	3.1	1558	17	T28255
C	38	46.6	3.1	1560	17	T18828
C	39	46.6	3.1	1582	17	T28259
C	40	46.6	3.1	1582	17	T18831
C	41	46.6	3.1	7063	20	X81916
C	42	46.6	3.1	7065	20	X81788
C	43	46.6	3.1	22481	17	T11658
C	44	46.4	3.0	2836	19	V40744
C	45	46.4	3.0	2836	19	V40745

ALIGNMENTS

RESULT	1
T37313	T37313 standard; cDNA to mRNA, 1518 BP.
XX	XX
AC	T37313:
XX	XX
DT	06-FEB-1997 (first entry)
XX	XX
DE	Aromatic acyl transferase coding sequence.
XX	XX
KW	Aromatic acyl transferase; transformation; anthocyanin pigment;
KW	plants; acylation; colour; tone; colouration; colour change;
KW	Gentiana triflora; Petunia hybrida; Petilla octimoides;
KW	Scenecio cruentus; Lavandula angustifolia; ds.
XX	XX
OS	Lavandula angustifolia (Clone PLAT21).
XX	XX
FE	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1355
FT	/*tag= a
FT	/product= Aromatic acyl transferase.
PN	XX
XX	WO9625500-A1.
XX	XX
PD	22-AUG-1996.
XX	XX
PE	16-FEB-1996; 96WO-JP00348.
XX	XX
PR	30-JAN-1996; 96JP-0046534.
PR	17-FEB-1995; 95JP-0067159.
PR	29-JUN-1995; 95JP-0196915.
XX	XX
PA	(SUNR) SUNTORY LTD.
XX	XX
PI	Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;

Malaria-specific D
Aromatic acyl tran
Human Immunodeficient
Clone am996-12 enc
cDNA encoding huma
Shuttle vector PMU
Sequence encoding
P.falciparum GBP13
Sequence encoding
Human osteopontin
Sequence encoding
Dictyostelium plas
Ligand-induced gen
Human foetal brain
Rhodospirillum rubrum
Human secreted pro
Actin gene terminal
Sequence encoding
Sequence of ANS-1
Candida boidinii A
Actin gene. Candi
Dictyostelium plas
3' fragment of clo
DNA encoding a hum
Survival motor neu
Human survival mot
Human survival mot
CDNA encoding huma
CDNA encoding huma
PEDF full length s
C. felis esterase,
C. felis esterase,

QY 1443 TTTTGGAGTCATTAATAAAAAAAAAAAAAAAAAATGAAA 1487
DB 1435 tggltcaataaagtgtaagcctcttgaataaaaaaaaaa 1479

RESULT 3
T37308
T37308 standard; cDNA to mRNA; 1703 BP.
T37308;
06-FEB-1997 (first entry)

DE Aromatic acyl transferase coding sequence.
XX
XX Aromatic acyl transferase; transformation; anthocyanin pigment;
KW plates; acylation; colour; tone; colouration; colour change;
KM Gentiana triflora; Petunia hybrida; Perilla ocimoides;
KM Senecio cruentus; Lavandula angustifolia; ds.
XX
XX Gentiana triflora var. japonica (clone pcAT4).
XX
FH Key Location/Qualifiers
FT CDS 6..1415
FT /tag= a
FT /product= Aromatic acyl transferase.
XX
XX W09625500-A1.
XX
PD 22-AUG-1996.
XX
PE 16-FEB-1996; 96WO-JP00348.
XX
XX 30-JAN-1996; 96JP-0046534.
PR 17-FEB-1995; 95JP-0067159.
PR 29-JUN-1995; 95JP-0196915.
XX
XX (SUNR) SUNTORY LTD.
XX
PI Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
PI Nakao M, Tanaka Y, Yonekura K;
XX
XX WPI: 1996-393401/39.
DR P-PSDB: W047722.
XX
XX DNA coding for aromatic acyl transferase - for transforming plants
PT which produce anthocyanin pigments and thus altering colour tone,
PT e.g. of flowers
XX
PS Claim 4; Page 53-57; 94pp; Japanese.
XX
XX Vectors containing DNA fragments encoding proteins of plant origin
CC with aromatic acyl transferase activity may be used to transform
CC plants which produce anthocyanin pigments. The aromatic acyl
CC transferase acylates the pigments in the flower resulting in colour
CC tone changes and allowing new colourations to be produced. Six
CC specific DNA sequences encoding aromatic acyl transferase from
CC different plants are described in T37308-T37313.
XX
XX Sequence 1703 BP; 512 A; 353 C; 356 G; 482 T; 0 other;

Query Match 12.9%; Score 196.2; DB 17; Length 1703;
Best Local Similarity 51.3%; Pred. No. 4; 1e-29;
Matches 686; Conservative 0; Mismatches 603; Indels 48; Gaps 8;

QY 46 AGGAGGCGTGAAGACATCCCGCTGACCTTTCGACATGACGCTGCTGCTTCAC 105
DB 69 aagaatgtcgaagtaacgtacacgtacattcttcgataccctcgtgtcgaactgaa 128
QY 106 CCCATGCTTCAGCTCTTCTTCTAAGAACTCCCTGTTCGAAACCCGCTTCGGAAC 165
DB 129 aagaatgtcgaagtaacgtacacgtacattcttcgataccctcgtgtcgaact 188

QY 166 GTGCTCCGAACCTCAACATCTTATCTTAACCTCAACACTTCTTCCCTTTCA 225
DB 189 gtatcccttaacttaagccctcttctctcactcaacttaaacactacccctcgttagc 248
QY 226 TGCATCTAATCTAACCTCTATCGCGGAGAAATGCGGAGTTCGGGTA-----TCAG 279
DB 249 ggaatttgttga tgcgcgaaccaatcgcgcgaattcgcgaattcgaactcccgltg 308
QY 280 AACGGTACTCGGTTTCTTTCACGATTAATGAGTCTGGAGAG---TCATCCGCAATTC 336
DB 309 gaaggcgaactcgaactaacttgcgttcgcgaatcgcgaactcgaactcgaactaa 368
QY 337 GCTCAATAATCTACTGCTTTCGCTTCCAGGAGATTAAGATCTCCAGTCCGCGC 396
DB 369 gtcatacaacttgtatagatccaatgtatgcagccttcttattgtatgcacaggt 428
QY 397 ATAGTGAGGAATCTGATCGGAATTTGTTCAAGTTTAAAGCTGCAAGTACTGTTT 456
DB 429 ataagaccatgcgaagacataaagtgatccgcgtcgcgtgcgaatgaaccglttt 488
QY 457 CCGGTCGGGGGTGTCATTCGAATACGACGACACCGCTTACGATGCTCCATG 516
DB 489 cctaaccgtggaatagccgtgcgtcgaagcgaatcgaatcgaatcgaatcgaat 548
QY 517 TTTGTAGGTTTATGAAGAGTTGGCTTCATCTAAATTCGAGAGATGATGAATTC 576
DB 549 ttgtatgttcaactcgtcgtgcgtatataaacaatttgcgaagaagcgcgtcgt 608
QY 577 TTGACGGAAGGTAATGTTTCCGCTTTCGACGATCGCTGTAATTTCCGCTT 636
DB 609 ttgtccgcga-----atcttctcactcttgcgaatagatcgaatcgaatcgaat 662
QY 637 AAATGACACATATTATGAAGAGGCGGACGACGACGACGACGACGACGACGACGAC 687
DB 663 ggcctagaggaacacatttgcgaagcgaagcgaagcgaagcgaagcgaagcgaag 722
QY 688 CAGCATCATCTTTTACCGAGCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 747
DB 723 ggaagcaaacccctcgtatccaagcgaagcgaagcgaagcgaagcgaagcgaag 782
QY 748 ATTAAAGAAATTGAAG-----GTTTGAATTCAGAGAAAGCCCAATGATGTAATTC 801
DB 783 attcagaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaag 842
QY 802 TCTTCTTGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 851
DB 843 acgaagcgtcaaatgagcgtgtgtgatacgtacgtacgtacgtacgtacgtacgtac 902
QY 852 --GGTCGAGATTACAAAGACGTGATGACGACAAACGCGCTTTTAAATTCGATTC 909
DB 903 gtgcatacgaaggaatcgtacgaagcgaagcgaagcgaagcgaagcgaagcgaag 962
QY 910 GATTTAAGGCGCGCTTTCGATTCGCGGCTCCGCGGAACTACTTCGGAATGCTATG 969
DB 963 gattgcggaagcgttgcgtaccccggtgtccgcaactacttgcgtacgttgcgt 1022
QY 970 TTTGCGATGCGGAAGATTCGCGGCGGATTTGCTGAGATGAAGGCTGTTCCGGA 1029
DB 1023 tcatgctgtcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaag 1082
QY 1030 GCTGAGCGATTCGCGGCGGAAATAGAGAAAGAGAGC--AGCGACAAAGATTCTAGA 1086
DB 1083 gtgcagcgtatgcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1142
QY 1087 ACTGTGGAAGATTCGCGGCTGAGATTCGCGGAAGCCTTCGAAATGTTATTTCCG 1146
DB 1143 gatgcgaagcgttgcgtacgtacgtacgtacgtacgtacgtacgtacgtacgtac 1202
QY 1147 GCGGATTCGAGAGGCTTATGATTCGCGGCGGATTTGAGTGGGTAAGCGCGGAG 1206
DB 1203 accgagtcgcttaagtcgtatcgtatgtgtgtatgttgcgtatgtgtgtgtgtgt 1262

Oy		1207	CAGAGATCTACTGTCATTGTATGGACGAAGTTTACCAGTGTCGTTGTAACCCGAGGAGT	1266
Dd		1263	tttgcattaccctcgttgtatgacgaagt--gatlatgtatccagtcaggat	1319
Oy		1267	GCGGCCGAGGAAWTTGGAGGTTGGATGCTGCTTGCCCAAAGAGAARTGCAGCTTTGGAT	1326
Dd		1320	tttgaaaaaaagggcgtagcatcggaagatcacatgacctaatcatatgatgatcatgca	1379
Oy		1327	GATTATTTTTCGCGAGCG	1343
Db		1380	aaaactcttgagaagaag	1396
RESULT	4	T37312		
ID	T37312	standard; cDNA to mRNA;	1508 BP.	
XX	XX	T37312:		
XX	DT	06-FEB-1997	(first entry)	
XX	XX	Aromatic acyl transferase coding sequence.		
Kw	XX	Aromatic acyl transferase; transformation; anthocyanin pigment;		
Kw	XX	plants; acylation; colour; tone; colouration; colour change;		
Kw	XX	Gencliana triflora; Petunia hybrida; Perilla ocimoides;		
Xx	OS	Senecio cruentus (Clone pCAT8).		
FH	Key	Location/Qualifiers		
FT	CDS	3..1367		
FT	/tag=	a		
FT	/product=	Aromatic acyl transferase.		
Pn	XX	M09625500-A1.		
PD	XX	22-AUG-1996.		
PF	XX	16-FEB-1996;	96MO-JP00348.	
PR	XX	30-JAN-1996;	96JP-0046534.	
PR	XX	17-FEB-1995;	95JP-0067159.	
PA	XX	29-JUN-1995;	95JP-0196915.	
XX	XX	(SUNR) SUNTORX LTD.		
PI	XX	Asahikari T., Fujiwara H., Fukui Y., Kusumi I., Mizutani M.;		
DR	XX	Nakao M., Tanaka Y., Yonekura K;		
XX	XX	WP1: 1996-393401/39.		
XX	XX	P-PSDB; W04726.		
PT	XX	DNA coding for aromatic acyl transferase - for transforming plants		
PT	XX	which produce anthocyanin pigments and thus altering colour tone,		
XX	XX	e.g. of flowers		
PS	XX	Claim 4: Page 69-72; 94pp; Japanese.		
CC	XX	Vectors containing DNA fragments encoding proteins of plant origin		
CC	XX	with aromatic acyl transferase activity may be used to transform		
CC	XX	plants which produce anthocyanin pigments. The aromatic acyl		
CC	XX	transferase acylates the pigments in the flower resulting in colour		
CC	XX	tone changes and allowing new colourations to be produced. Six		
CC	XX	specific DNA sequences encoding aromatic acyl transferase from		
XX	XX	different plants are described in T37308-T37313.		
SQ	Sequence	1508 BP; 442 A; 294 C; 295 G; 477 T; 0 other;		
Query Match	11.4%;	Score 174.2;	DB 17;	Length 1508;
Best Local Similarity	48.3%;	Pred. No. 6.4e-25;		
Matches 730; Conservative	0;	Mismatches 738; Indels	43; Gaps	7

[illegible]


```

Db 710 gaaagaaactgctcaaatctcaggcaccactactgttcttaatccagcaatttctaa 769
QY 709 GATGGGATTTGAGTACTTACTTTCACCCAACTGAAATTAAGAAATTGAAGGTTTG 768
XX ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 770 gatgaagltcgagccacttcatccatctacacccatctgatcagaagctcaagaatttc 829
QY 769 ATTGAGAGAAAAGCCCAATGTGTA-----GTTAATCTCTCTTCCTGTC 813
XX ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 830 atttcgcataaaaaatcgcaacttaacccgtagtagtaataataatcgtcaacttcaag 889
QY 814 GCGATCCAGCTTATCTGAGCCGCATCCCAAT-----CGTCCGAGATTAAC 864
Db 890 gtgacatctgacactgacatctgacatctgacatctgacatctgacatctgacatctgac 949
QY 865 AAAACAGCTGATGACGACAAACGGCTTCTTTTAAATTCGATGATTTAAGCCCGCT 924
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 950 aaggctgagaagagataaacaactgacaaactatgcttcatcaactcgcgcaacgct 1009
QY 925 TTGGATCCGCGCGCTCCGGGAACTACTTTCGAAACTGATCTTATCGTTTGGATGCGAAG 984
Db 1010 ttgtctcgcgcgataactcaaaatcttggaaatcgatagatgcttctatgtagtga 1069
QY 985 ATCTCGCGCGGATTTGGTGGAGATGAAGGGGTGTTGCGGACGCTGACGCGATCGCG 1044
Db 1070 tcgactcatgagcaactgttagaagaggtgtgctgagtagctgcaacgcatcgga 1129
QY 1045 GCGGAATATAGAGAGAGAGACGACAGAGAA---GATTCAGAACTGTGGAAGACTGG 1101
Db 1130 gatgctatccataagaaggttatactgactacgaagaattctgagaagagatgatactg 1189
QY 1102 CCGCTGAGATTCGCGAAGCCTTGCAAAACT-----GTTATTTCTCGTGGCGGGA 1152
Db 1190 cgcgcgcgcgatacaaatctgctgcgcaacgctcgaagctcattatctgctgtagctcga 1249
QY 1153 TCAGACAGGCTTGTATCTTTACGCGCGGATTTTGATGAGGAGTGAACAGAG 1212
Db 1250 caacgcaatgcatgcatgatttgcagatcttgcgtggaagagcttgaagaacatgaa 1309
QY 1213 ATACTGCGATGTATGAGAGAGACTTTACGATGCTGTTGTAACCGAGGATGCTGCC 1272
Db 1310 tctgttcaactaaacctctcgcgcaacataatttgcctcctcgcgacgaagaatttaa 1369
QY 1273 GGAGGATTTGGAGTTGATGCTTTTGCAAAAGAGAGAAATGCAAGCTTTGATGATTAAT 1332
Db 1370 ggaagcaactgagcttgcacatcttgcctaagaatagatgagcatttgcacat 1429
QY 1333 TTTCGAGGAGGATTAAGGAT 1353
Db 1430 ttacgaattcattcaatagtc 1450

```

```

PR 27-MAR-1998: 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
PI Rosenthal A, Rump A, Hess J, Aligner T, Wirth T;
DR WPI; 1999-601320/51.
XX
PS
PT Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX
XX Example 3: Page 300-328; 391pp; German.
XX
CC This invention describes novel nucleic acids (i: designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes a human LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 12459 A; 12933 C; 12356 G; 12251 T; 0 other:

Query Match 4.1%; Score 62.4; DB 20; Length 49999;
Best Local Similarity 68.0%; Pred. No. 0.0023;
Matches 87; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1395 CTTCTGTTTATCTCTATGTTTAACAAATTAATTTTTCATGAACTTTTGGACTCA 1454
Db 41887 cctccctcctctcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 41946
QY 1455 ATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1514
Db 41947 tttttaaataaataaataaataaataaataaataaataaataaataaataaataaataa 42006
QY 1515 TTTTTTYT 1522
Db 42007 tttttttt 42014

RESULT 7
223902/C
ID 223902 standard; DNA; 49999 BP.
XX
AC 223902;
XX
DE 25-JAN-2000 (first entry)
XX
XX Human LOBO homologue genomic DNA fragment 4.
XX
XX OS Homo sapiens.
XX
XX PN W09950284-A2.
XX
XX PD 07-OCT-1999.
XX
XX PR 26-MAR-1999; 99WO-EP02055.
XX
XX PA (ROSE/) ROSENTHAL A.
XX
XX PI Rosenthal A, Rump A, Hess J, Aligner T, Wirth T;

```


DR	WPI: 1999-601320/51.	
XX		
PT	Nucleic acids encoding proteins which influence bone development,	
PT	useful for treating and studying bone disorders -	
XX		
PS	Example 3: Page 300-328; 391pp: German.	
XX		
CC	This invention describes novel nucleic acids (1: designated LOBO (long	
CC	bones)) encoding proteins influencing bone development in mammals. The	
CC	proteins of the invention reduce and/or inactivate bone extension (i.e.,	
CC	development), with exception of the skull and have osteopathic activity.	
CC	The nucleic acid molecules, proteins and antibodies can be used in	
CC	diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods	
CC	and nucleic acid molecules, etc. are useful for production of transgenic	
CC	animals, especially a transgenic mouse for the study of diseases	
CC	associated with bone development, e.g. spondyloepiphyseal dysplasia and	
CC	achondroplasia. This sequence encodes a human LOBO protein described	
CC	in the method of the invention.	
SO	Sequence 49999 BP: 12459 A; 12933 C; 12356 G; 12251 T; 0 other;	
Query Match	3.6%; Score 55.4; DB 20; Length 49999;	
Best Local Similarity	84.9%; Pred. No. 0.049; Mismatches 11; Indels 0; Gaps 0	
Matches 62; Conservative 0; Mismatches 11; Indels 0; Gaps 0		
OY	1450 AGCTCAATGAAAAAAAAAAAAAAAAAATGAAAAAAAACTCAGTTATTTTATTTT 1509	
Db	42026 AGNAAATTTTATTTT 41967	
OY	1510 TTTTATTTTATTTT 1522	
Db	41966 TTTTATTTTATTTT 41954	
RESULT 8		
Q27886/C		
ID	Q27886 standard; DNA: 2418 BP.	
XX		
AC	Q27886;	
XX		
DT	04-FEB-1993 (first entry)	
XX		
DE	P.falciparum GBP130h.	
XX		
KW	Polymerase chain reaction; glycoprotein binding protein; inverse PCR;	
KW	Plasmodium falciparum; malaria; ss.	
XX		
OS	Synthetic.	
XX		
Key	Location/Qualifiers	
FT	exon	767..955
FT	/*tag= a	/number= 1
FT	exon	1111..2202
FT	/*tag= b	/number= 2
FT	repeat_region	1249..2202
FT	/*tag= c	
XX		
PN	EP499834-A.	
XX		
PD	26-AUG-1992.	
XX		
PE	27-JAN-1992; 92EP-0101271.	
XX		
PR	21-FEB-1991; 91DE-4105348.	
XX		
PA	(BEHM) BEHRINGWERKE AG.	
XX		
PI	Hundt E, Knapp B, Kupper H, Nolte D, Kuepper H;	
XX	WPI: 1992-286009/35.	
DR		

XX	PT	DNA coding protein GBP_130 h and proteins produced - used for vaccines to control malaria	XX
XX	PS	Claim 1; Page 12-14; 17pp; German.	XX
XX	CC	Glycophorin binding protein 130h is a P.falciparum blood stage antigen 69k homologous with the known GBP130. The GBP130h gene and its homologues can be used in vaccines to protect against malaria caused by P.falciparum. The coding sequence was isolated using PCR techniques on a genomic P.falciparum FC8R strain library. The GBP130h gene was found to be highly conserved between different CC strains but is distinct from the GBP130 gene.	XX
XX	CC	See Q27878-Q27885 and Q33432.	XX
XX	CC	Sequence 2418 BP; 1047 A; 303 C; 292 G; 776 T; 0 other:	XX
XX	CC	Query Match 3.6%; Score 55; DB 13; Length 2418; Best Local Similarity 55.5%; Pred. No. 0.038; Matches 106; Conservative 0; Mismatches 85; Indels 0; Gaps 0	XX
XX	CC	1332 TTTTCGAGGAGAAATGAAGGTTGATTATCATTTAATCATCATTAATGAGTGGATGA 1391	XX
XX	CC	766 TTTTACTTAGAAGAAATAAGAAATATATATGTTTTCTCAATTTGCTGCTCTAAATA 707	XX
XX	CC	1332 AATCCTCGTTCATCTCATTTGTTAAACAATTAATTTTTTCCATGTGACCTTTTGGAG 1451	XX
XX	CC	706 ATTATATAATATATAATATATATATATATATATTTTTTTTAAATATATATTTTAA 647	XX
XX	CC	1452 TCAATATAAAAAAAAAAAAAAAAAAAAAAGAAAAAACTCAGTATTTTTTTTTTTTT 1511	XX
XX	CC	646 ATATAATAATATAATAATAATAATAATAACAAATACATTAATTTTTTTTATTTATTT 587	XX
XX	CC	1512 TTTTTTTTTT 1522	XX
XX	CC	586 TAAATTTTTTTTT 576	XX
XX	CC	RESULT 9	XX
XX	CC	N90703/c	XX
XX	CC	N90703 standard; DNA: 2662 BP.	XX
XX	CC	N90703;	XX
XX	CC	25-JAN-1990 (first entry)	XX
XX	CC	Rhoptry membrane antigen-1 gene.	XX
XX	CC	Malaria; rhoptry membrane antigen-1; antibodies.	XX
XX	CC	Plasmodium falciparum D10.	XX
XX	CC	Key Location/Qualifiers	XX
XX	CC	CDS 634..2202	XX
XX	CC	/*tag= a	XX
XX	CC	W08907645-A.	XX
XX	CC	24-AUG-1989.	XX
XX	CC	10-FEB-1989; 89WO-AU00056.	XX
XX	CC	12-FEB-1988; 88AU-0006743.	XX
XX	CC	(SARA) SARAWANE PTY LTD.	XX
XX	CC	Peterson MG, Crewther PE, Smythe JA, Marshall VM, Silva A;	XX
XX	CC	WPI; 1989-263714/36.	XX
XX	CC	P-PSDB; P91632.	XX
XX	CC	Rhoptry membrane antigen of Plasmodium falciparum	XX

PT - used for producing antibodies and in immunisation,
 PT diagnostic and treatment methods for malaria.
 PS Claim 8; fig 1; 46pp; English.

CC The RMA-1 gene can generate an immune response to malaria, and
 CC antibodies which can inhibit growth of the parasite. RMA-1 initially
 CC has mol. wt. 80 kD.
 XX

SO Sequence 2662 BP; 1134 A; 307 C; 387 G; 834 T; 0 other;

Query Match
 Best Local Similarity 3.5%; Score 54; DB 10; Length 2662;
 Matches 84; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 1388 ATGAATCCCTGTTTCATCTCTATTTTAAACATATTTTTCATTTGAACCTTTT 1447
 DB 288 ATAAATTTATAGCTTTTAAATTTATTTAAATATAATATATTTTCAATTTTATCT 229
 OY 1448 TGAGTCATTAATAAAAAAAAAAAAAAAAAATGAATAACCTCATTTTCTTTT 1507
 DB 228 CATTTTCTTTTAAATAAAAAAAAAATTAACCATTAATATGTTAATTTTCTT 169
 OY 1508 TTTTCTTTTCTTTT 1521
 DB 168 TTTTCTTTTCTTTT 155

RESULT 10
 T48669
 ID T48669 standard; cDNA; 1875 BP.
 AC T48669;

DT 29-APR-1997 (first entry)

DE Human EDG-2 receptor homologue (HEDG) cDNA sequence.

KW EDG-2; Rheumatoid synovium; G-protein linked; seven pass; infection;
 KW bacterial; viral; fungal; allergy; lymphoma; carcinoma; cancer;
 KW neoplasia; lymphocyte; leucocyte trafficking; ss.

OS Homo sapiens.

PH Key Location/Qualifiers
 FT 309..1403
 FT CDS /*tag- a

PN MO9700952-A2.

PD 09-JAN-1997.

PE 20-JUN-1996; 96WO-US10618.

PR 06-DEC-1995; 95US-0567817.

PR 20-JUN-1995; 95US-0000352.

PA (INCY-) INCYTE PHARM INC.

PI Au-Young J, Bandman O, Coleman R, Guegler KJ, Seilhamer JJ;

DR WPI; 1997-087379/08.

DR P-PSDB; W10879.

PT New isolated human EDG-2 receptor homologue - used to develop prods.
 PT for the diagnosis or treatment of inflammation or diseases which
 PT activate genes of lymphoid tissues.
 PS Claim 2; Page 25; 35pp; English.

CC T48669 is a cDNA sequence encoding the human EDG-2 receptor
 CC homologue, HEDG. The cDNA sequence can be used for the production of

CC primers and probes which may be used in the diagnosis of viral,
 CC bacterial or fungal infections, allergic responses; hereditary,
 CC disease, lymphoma or carcinoma or other conditions which activate
 CC the genes of the lymphoid tissue. Inhibitors of HEDG, e.g. HEDG-
 CC specific antibodies, can be used to treat inflammation or disease as
 CC described above. Inhibitors of HEDG are also useful for controlling
 CC signal transduction in rheumatoid synovium (from where it is derived)
 XX

SO Sequence 1875 BP; 478 A; 477 C; 410 G; 510 T; 0 other;

Query Match
 Best Local Similarity 3.5%; Score 52.6; DB 18; Length 1875;
 Matches 97; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

OY 1319 CTTTGATGATTTATTTGCGAGGGAATTAAGCTTGATTAATCATTTATCATTTA 1378
 DB 1703 ctttaaaagattcttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1762
 OY 1379 TGAAGTTGATGAATTCCTCTGTTTCATCTATTTGTTAAACATTAATTTTTCATT 1438
 DB 1763 acaacttcattatatacaaggctcccttcttattttaaaggatacgcttcactaat 1822
 OY 1439 GAACCTTTTTCATTAATAAAAAAAAAAAAAAAAAATGATAAAAA 1489
 DB 1823 aaacacgttattcctctaaataaaaaaaaaaaaaaaaaaaaaaaaaa 1873

RESULT 11
 Q11712/c
 ID Q11712 standard; DNA; 3138 BP.
 AC Q11712;

DT 30-JUL-1991 (first entry)

DE Shuttle vector PMUW1630.

KW slime mould; replication; Rep gene; Ddp2; ss.

OS Dictyostelium discoideum.

PH Key Location/Qualifiers
 FT sig-peptide 2068
 FT /*tag- a
 FT /note- "finish position not given"

PN MO9106644-A.

PD 16-MAY-1991.

PE 02-NOV-1990; 90WO-AU00530.

PR 02-NOV-1989; 89AU-0007187.

PA (UYMA-) MACQUARIE UNIV.

PI Slade MB, Chang ACM, Williams KL;

DR WPI; 1991-164194/22.

DR P-PSDB; W10879.

PT Polypeptide facilitating extra-chromosomal replication - of
 PT recombinant plasmid in Dictyostelium species
 PS Claim 35; Fig 15; 90pp; English.

CC Shuttle vector PMUW1630 was constructed for use in E.coli and
 CC transfer back into Dictyostelium. It includes a fragment of the
 CC Dictyostelium plasmid Ddp2 containing the origin of replication,
 CC the promoter from the Dictyostelium Actin 15 gene, a sequence
 CC encoding the secretion signal peptide of the D19 gene of protein
 CC PSA. Actin 15 polyA signal and an ampicillin resistance-conferring
 CC sequence. See also Q11710 and Q11711.

Db 4982 aatttaatgccaaaagtttgccatgcttaacacataactataatttccctta 5041
Qy 1388 ATGAATCCCTGCTTCATCTCATTTTAAACAATTAATTTTTCATGAACTTTT 1447
Db 5042 ataaacacttttgttaaatgtatcttcctttaataaataatttaagcaatgtcca 5101
Qy 1448 TGAGTCATTAATAAAAAAAAAAAAAAAAAATGAAAAA 1489
Db 5102 ataaaaaaaaaaaaaaaaaaaaaaaaaaaaa 5143

Search completed: October 20, 2000, 03:16:45
Job time: 5992 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 03:11:44 ; Search time 103.47 Seconds
(without alignments)
2224.698 Million cell updates/sec

Title: US-08-894-356C-6
Perfect score: 1522
Sequence: 1 NFGACCACTCTCGAATC.....TTTTTTTTTTTTTTTTTTT 1522

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/6_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PCITUS_COMB.seq: *
7: /cgn2_6/ptodata/1/ina/backfillseq.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	3.7	7218	1	US-08-232-463-14
2	56.8	3.7	7218	1	US-08-232-463-14
3	52.6	3.5	1875	6	PCT-US96-10618-1
4	51.4	3.4	3138	1	US-07-867-106-4
5	50	3.3	5173	1	US-08-242-677-1
6	49.8	3.3	130	7	5198345-15
7	49.8	3.3	130	7	5198345-15
8	49.6	3.3	3138	1	US-07-867-106-4
9	49.6	3.3	5852	1	US-07-867-106-2
10	48.4	3.2	1493	7	5340934-5
11	47.6	3.1	198	6	US-08-330-108-15
12	47.6	3.1	198	6	PCT-US92-10087-16
13	47.2	3.1	5852	1	US-07-867-106-2
14	46.8	3.1	860	2	US-07-847-010-13
15	46.8	3.1	4818	5	US-08-817-926-27
16	46.6	3.1	1582	5	US-08-545-1968-10
17	46.6	3.1	1582	5	US-08-545-1968-12
18	46.6	3.1	22481	6	PCT-US95-07201-43
19	46.4	3.0	2836	5	US-08-747-221B-34
20	45.6	3.0	3095	7	US-08-747-221B-36
21	45.6	3.0	3095	7	5231168-1
22	45.4	3.0	2209	2	US-08-514-014-1
23	45.4	3.0	2209	4	US-08-833-823-1
24	45.4	3.0	2280	5	US-08-813-150-1
25	44.6	2.9	860	1	US-07-847-010-18
26	44.6	2.9	2422	1	US-07-867-106-5

27	43.8	2.9	222	6	PCT-US93-00869-15	Sequence 15, Appl
28	43.6	2.9	83	1	US-08-420-443-5	Sequence 5, Appl
29	43.6	2.9	919	4	US-08-924-759-19	Sequence 19, Appl
30	43.6	2.9	919	5	US-09-248-335-19	Sequence 19, Appl
31	43.6	2.9	4507	3	US-08-568-459A-3	Sequence 3, Appl
32	43.6	2.9	4507	4	US-08-487-826B-3	Sequence 3, Appl
33	43.4	2.9	1931	4	US-09-019-942-2	Sequence 2, Appl
34	43.4	2.9	2502	5	US-09-234-332-1	Sequence 3, Appl
35	43.2	2.8	2058	4	US-08-749-331-1	Sequence 1, Appl
36	42.8	2.8	665	4	US-08-883-795A-36	Sequence 1, Appl
37	42.8	2.8	2058	4	US-08-749-331-1	Sequence 36, Appl
38	42.6	2.8	319	1	US-07-593-657-14	Sequence 1, Appl
39	42.6	2.8	946	5	US-08-916-443A-15	Sequence 15, Appl
40	42.6	2.8	2007	5	US-08-747-221B-36	Sequence 36, Appl
41	42.6	2.8	2007	5	US-08-747-221B-38	Sequence 38, Appl
42	42.4	2.8	1532	1	US-08-248-466B-11	Sequence 11, Appl
43	42.4	2.8	2897	4	US-08-927-394-1	Sequence 1, Appl
44	42	2.8	2781	5	US-08-749-522-4	Sequence 4, Appl
45	42	2.8	3095	7	5231168-1	Patent No. 5231168

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEITLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PL-F15
US-08-232-463-14

Query Match 3.7%; Score 57; DB 1; Length 7218;
Best Local Similarity 4.1%; Pred. No. 0.00028;
Matches 15; Conservative 210; Mismatches 140; Indels 0; Gaps 0;

QY 5 CCACCTCTCCGATCTCCCGAGTGGCCGCTCCAGGACGGTGGCTGACGACTCAG 64
DB 1099 YY 1158
QY 65 TCCCGCTCACCTCTCGACATGAGCTGGCTGATTCACCCGACGCTCAGCTCTCT 124
DB 1159 YY 1218
QY 125 TCTACGACCTCCCTGCTCCAAACCCGCTCTCGAAACCGTCTCGAACTCAAC 184
DB 1219 YY 1278
QY 185 AATCTTATCTCAACCTCAACACTTCTCCCTTCATGCAATCAATCTACCTC 244
DB 1279 YY 1338
QY 245 TATCGCGGAGAAATGCGGAGTTCGGATTCAGAAAGGTGATCGGTTCTTCACGA 304
DB 1339 YY 1398
QY 305 TTAGGAGTCTGTGAGATCATCGCATTCGCTCATAAATCTACTGCTTGCCCTA 364
DB 1399 YYGTACAAATCTCTATCTCTTA 1458
QY 365 GGGAC 369
DB 1459 ACTAC 1463

RESULT 2
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300, 6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PT2gpt-F1s
US-08-232-463-14

Query Match 3.7%; Score 56.8; DB 1; Length 7218;
Best Local Similarity 4.0%; Pred. No. 0.00031;
Matches 16; Conservative 225; Mismatches 157; Indels 0; Gaps 0;

QY 967 TCGTTGCCATGCGCAAGATCTCGCGCGGATTTGTCGAGATCAAGCGCTTCG 1026
DB 1462 TAGTTAAAGATGAGATGATTTGTACRRRRRRRRRRRRRRRRRRRRRRRR 1403
QY 1027 GCAGCTGAGCGCATCGCGCGGAATAGAGAGAGCAGCAGCAAGATTTAGAA 1086
DB 1402 RRR 1343
QY 1087 ACTGTGAGAACTGCGCTGTGAGATTCGCAAGCCTTGCAAACTGTAATTCGTG 1146
DB 1342 RRR 1283
QY 1147 GCGGCTGAGCGAGCTGATCTTACGGCGGATTTGGATGGGTAGCGGTGAG 1206
DB 1282 RRR 1223
QY 1207 CAAGAGATCTGTGATGATGAGAGATTACGATGCTGTGTAACGAGGAT 1266
DB 1222 RRR 1163
QY 1267 GCTGCCGAGATGAGGTTGGATTGCTTTGCCAAGAGAGATTCAGCTTTGAT 1326
DB 1162 RRR 1103
QY 1327 GATTATTTCGCGAGGATTAAGGTTGATTAATAT 1364
DB 1102 RRR 1065

RESULT 3
PCT-US96-10618-1
Sequence 1, Application PC/TUS9610618
GENERAL INFORMATION:
APPLICANT: Coleman, Roger
APPLICANT: Guegler, Karl J.
APPLICANT: Au-Young, Janice
APPLICANT: Bandman, Olga
APPLICANT: Sellhammer, Jeffrey J.
TITLE OF INVENTION: A NOVEL HUMAN EDG-2 RECEPTOR HOMOLOG
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10618
FILING DATE: 20-JUN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/000,352

RESULT 4
 US-07-867-106-4/C
 ; Sequence 4, Application US/07867106
 ; Patent No. 5389526
 ;
 GENERAL INFORMATION:
 APPLICANT: Slade, Martin B
 APPLICANT: Chang, Andy C M
 APPLICANT: Williams, Keith L
 TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
 TITLE OF INVENTION: Slime Moulds of the genus Dictyostelium
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526R1S
 STREET: One Liberty Place 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/867,106
 FILING DATE: 19920625
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU PJ 7187
 APPLICATION NUMBER: PCT/AU90/00530
 FILING DATE: 02-NOV-1989
 ATTORNEY/AGENT INFORMATION:

```

US-08-242-677-1
; Sequence 1, Application US/08242677
; Patent No. 5677143
; GENERAL INFORMATION:
; APPLICANT: Gaynor, Richard B
; APPLICANT: Wu, Poon W.
; TITLE OF INVENTION: Cellular Nucleic Acid Binding Protein
; TITLE OF INVENTION: and Uses Thereof in Regulating Gene Expression and in the
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/242,677
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mayfield, Denise L.
; REGISTRATION NUMBER: 33,732
; REFERENCE/DOCKET NUMBER: UTSD:401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```



```

1 GENERAL INFORMATION:
2 APPLICANT: Slade, Martin B
3 APPLICANT: Chang, Andy C M
4 APPLICANT: Williams, Keith L
5 TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
6 TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
7 NUMBER OF SEQUENCES: 19
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rls
10 STREET: One Liberty Place 46th Floor
11 CITY: Philadelphia
12 STATE: PA
13 COUNTRY: USA
14 ZIP: 19103
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/07/867,106
22 FILING DATE: 19920625
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: AU PJ 7187
25 APPLICATION NUMBER: PCT/AU90/00530
26 FILING DATE: 02-NOV-1989
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Feeney, Joanne Longo
29 REGISTRATION NUMBER: 35,134
30 REFERENCE/DOCKET NUMBER: RICE-0002
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: 215-568-3100
33 TELEFAX: 215-568-3439
34 INFORMATION FOR SEQ ID NO: 2:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 5852 base pairs
37 TYPE: NUCLEIC ACID
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: DNA (genomic)
41 ANTI-SENSE: NO
42 FEATURE:
43 NAME/KEY: CDS
44 LOCATION: 2378..5038
45 FEATURE:
46 NAME/KEY: CDS
47 LOCATION: 2378..5038
48 US-07-867-106-2
49
50 Query Match: 3.3%, Score 49.6; DB 1; Length 5852;
51 Best Local Similarity 69.8%; Pred. No. 0.013;
52 Matches 67; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
53
54 QY 1427 TTTTTCATCGAAGCTTTTTGACTCAATAAAAAAAAAAAAAAAATGCAA 1486
55 ||||| | ||||| || |||||||||||||||| || ||||
56 Db 2032 TTTTTCATTTTTTTTTTTTTCACAAATAAAAAAAAAAAAAAGAAATGAAA 1973
57
58 QY 1487 AAACACTGTATTATTTTTTTTTTTTTTTTTTTTTTTT 1522
59 || | |||| | || ||||| | ||||
60 Db 1972 AAAAGTTGGTTAAACTACATTTAGTTTATAGTTT 1937
61
62 RESULT 10
63 5340934-5
64 ; Patent No. 5340934
65 ; APPLICANT: TERMINE, JOHN D.; YOUNG, MARIAN F.; FISHER, LARRY W.
66 ; ROBEY, PAMELA G.
67 ; TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS
68 ; NUMBER OF SEQUENCES: 13
69 ; CURRENT APPLICATION DATA:
70 ; APPLICATION NUMBER: US/07/432,044
71 ; FILING DATE: 03-NOV-1989

```

```

; SEQ ID NO:5:
; LENGTH: 1493
5340934-5

Query Match          3.2%; Score 48.4; DB 7: Length 1493;
Best Local Similarity 64.0%; Pred. No. 0.016;
Matches 73; Conservative 0; Mismatches 41; Indels 0; Gaps 0

Qy 1376 TTTATGAAGTTCGATGAATCCCTCTGTTTCATCTCATGTGTTAAACAATAATTGTTTCC 1435
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1371 ttcttaagtagtgatatacttctgttgtgattcctttgcgtgacaaactctttac 1430

Qy 1436 ATTGAACCTTTTGGACTAATAAAAAAAATGCAAAAAA 1489
      ||||| + | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1431 cttagaatgctaagaagaaaaaaaataaaaaaaaaaaaaa 1484

RESULT 11
US-08-330-108-16
; Sequence 16, Application US/08330108
; Patent No. 5795752
; GENERAL INFORMATION:
; APPLICANT: Smith, Kendall A.
; TITLE OF INVENTION: IL-2-Stimulated Gene
; TITLE OF INVENTION: Expression
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahnive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/330,108
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/104,736
; FILING DATE:
; APPLICATION NUMBER: US/07/796,066
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; CELL TYPE: T-cell blast
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: 8DA-T3
; US-08-330-108-16

Query Match          3.1%; Score 47.6; DB 2: Length 198;
Best Local Similarity 61.1%; Pred. No. 0.012;
Matches 77; Conservative 0; Mismatches 49; Indels 0; Gaps 0

Qy 1364 TTTTAATCATGTTATATGAAAGTTGCGTAATCCCTGTTTCATCTCATGTGTTAAACAA 1423
      ||||| + | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 TTTATTTTGGAATTAAGTCGTTGCCCTTTATTTGTAAGCTGTTATTAATATATATTTATAT 127
```


Sun Oct 22 12:46:02 2000

us-08-894-356c-6.rni

Page 8

OY 1430 TTTTCCTGAGCACTTTTATGACATAAAAAAAAAAAAAAATGGAAA 1489
Db 4440 ATTAAACAACRAATTAAACATGATCAAAAAAAAAAATTAAACAAGATTAAACACAAA 4381

OY 1490 CTCAGTATTTTTTTTTTTTTTTTTTTTTT 1519
Db 4380 AGAACACGAATTTTTCTTTTATTTTAT 4351

Search completed: October 20, 2000, 03:12:31
Job time: 5769 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 02:07:23 ; Search time 1792.79 Seconds
(without alignments)
5248.931 Million cell updates/sec

Title: US-08-894-356c-6
Perfect score: 1522
Sequence: 1 NTGACCACCCCTCTCGAATC.....TTTTTTTTTTTTTTTTTTT 1522

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estln1:*
64: em_estln2:*
65: em_estln3:*
66: em_estln4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estpl1:*
71: em_estpl2:*
72: em_estpl3:*
73: em_estpl4:*
74: em_estpl5:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

117: gb-ss18:*
 118: gb-ss19:*
 119: em-ss13:*
 120: gb-ss20:*
 121: gb-ss21:*
 122: gb-ss22:*
 123: gb-ss23:*
 124: gb-ss24:*
 125: em-ss14:*
 126: em-ss15:*
 127: em-ss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146.6	9.6	606	20	AM221049 EST297518
2	138.6	9.1	583	20	AM221050 EST297519
3	121.8	8.0	687	20	AM102336 s886d06.y
4	116	7.6	591	35	BE434257 EST405335
5	112	7.4	518	23	AM650280 EST328734
6	108	7.1	487	23	AM616206 EST307245
7	86.4	5.7	497	25	AM930934 EST356777
8	84.8	5.6	490	19	AM038515 EST280198
9	83.2	5.5	409	19	AM093479 EST286659
10	81.8	5.4	788	34	BE187619 EST336180
11	79.2	5.2	597	24	AM774948 EST334099
12	79	5.2	656	24	AM692527 NF056D055
13	79	5.2	677	23	AM586972 EST318595
14	77	5.1	311	35	BE353747 EST355090
15	73.4	4.8	605	18	AV440958 AV440958
16	71.6	4.7	1101	121	CNS00293 BE323055 NF001A07P
17	70.2	4.6	668	34	BE323055 NF001A07P
18	69.8	4.6	694	23	AM560686 EST315734
19	67.4	4.4	581	122	CNS034DK AI227297 Tetradon
20	67	4.4	1146	122	CNS021G2 AI176843 Tetradon
21	65.6	4.3	587	123	CNS0464P AI276226 Tetradon
22	65.4	4.3	658	24	AM695291 NF033E09S
23	65.4	4.3	1062	121	CNS00CAA AL058926 Drosophila
24	65.2	4.3	1101	121	CNS0021T AL061936 Drosophila
25	65	4.3	1101	121	CNS00238 AI097166 Drosophila
26	65	4.3	1101	121	CNS00ZAL AI097431 Drosophila
27	64.8	4.3	1027	122	CNS02T50 AI212733 Tetradon
28	64.4	4.2	603	23	AM616119 EST296884
29	64.2	4.2	1201	121	CNS015WP AI106003 Drosophila
30	63.6	4.2	615	23	AM587016 EST318639
31	63.6	4.2	1092	122	CNS020K7 AI175696 Tetradon
32	63.2	4.2	1039	122	CNS02ADN AI188440 Tetradon
33	63.2	4.2	1101	121	CNS003BD AL064091 Drosophila
34	63	4.1	836	122	CNS02M02 AI213739 Tetradon
35	63	4.1	905	121	CNS00KH2 AI077798 Drosophila
36	63	4.1	928	121	CNS00DKY AI071865 Drosophila
37	63	4.1	1101	121	CNS003BD AI064091 Drosophila
38	63	4.1	1101	121	CNS0160C AI106998 Drosophila
39	62.6	4.1	987	121	CNS014PQ AI104456 Drosophila
40	62.4	4.1	456	123	CNS04MB2 AI297224 Tetradon
41	61.8	4.1	940	123	CNS03004 AI255325 Tetradon
42	61.6	4.0	572	9	AI297571 LP1102.5
43	61.6	4.0	1074	121	CNS015ZR AI106113 Drosophila
44	61.6	4.0	1101	121	CNS0021T AI061936 Drosophila
45	61.6	4.0	1101	121	CNS0100X AI098379 Drosophila

ALIGNMENTS

RESULT 1

LOCUS	AM221049	606 bp	MRNA	EST	07-DEC-1999
DEFINITION	EST297518	tomato fruit mature green, TAMU	Lycopersicon esculentum		
ACCESSION	CDNA clone CLEP3F1	RNA sequence.			
VERSION	AM221049				
KEYWORDS	AM221049.1	GI:6532733			
SOURCE	EST.				
ORGANISM	Lycopersicon esculentum				
REFERENCE	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Rønning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.				
AUTHORS	1 (bases 1 to 606)				
TITLE	Unpublished (1999)				
JOURNAL	Contact: David Fritsch				
COMMENT	Clemson University Genomics Institute				
	100 Jordan Hall, Clemson, SC 29634, USA				
	Tel: 864 656 4366				
	Fax: 864 656 4293				
	Email: dfritsch@CLEMSON.EDU				
FEATURES	5 prime sequence.				
source	Location/Qualifiers				
	1..606				
	/organism="Lycopersicon esculentum"				
	/cultivar="TA496"				
	/db_xref="taxon:4081"				
	/clone="CLEP3F1"				
	/clone_lib="tomato fruit mature green, TAMU"				
	/tissue_type="fruit pericarp"				
	/dev_stage="mature green (3-5 days pre-ripening)"				
	/lab_host="SOLR"				
	/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2: XhoI; CLEP - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"				
BASE COUNT	147 a 143 c 123 g 193 t				
ORIGIN					
Query Match	9.6%; Score 146.6; DB 20; Length 606;				
Best Local Similarity	53.4%; Pred. No. 1.7e-20;				
Matches 308; Conservative	0; Mismatches 269; Indels 0; Gaps 0;				
QY	2 TGACCAACCCCTCGAATCCCTCCGAGTGGCGCCCTCCAGCAGCGTGCAGCAGT	61			
DB	5 TGGCCACCGGATGTGACATGTGCAAGTTCGCCACTCCGGCGGCGAAGGAGTGA	64			
QY	62 CACTCCCGCTCACCTTCGACAGTGGCTGCATTTCCACCCCATGCTTCACTTC	121			
DB	65 TACTCCCTTACTTATTTGACCATGTTTGTTAGGAGTTCCCGCATGAGCGGATAT	124			
QY	122 TCTTACGACATCCCTGTTCGAACCGGCTTCTCGAAGCGTTCGCAAACTCA	181			
DB	125 TATTTTACAAGCTCTCCATTTTCAACCCGATTTTCGTTCAAACTATATCCCTCTTA	184			
QY	182 AACAACTTATCTTAACCTCAACACTCTTCCCTTCATGCAATCTAATCTAC	241			
DB	185 AAAATTCACCTCTCTCACTCAACACTATACGCGCTTAGCGGAACGTTGCTTGC	244			
QY	242 CTCATCGCCGGAGAAATGCCGAGTCCGATATGAGAACGCTGCTTCTTTCA	301			
DB	245 CACTAGTACAAAGGATATCTCGAGTACGTTATGTGACAGAGATTCGTCTGTTA	304			
QY	302 CGATTATGAGCTGTTCGAGATCATCCGATTCGCTCAATAATCTACTGTTGCC	361			
DB	305 CTTTTCGAGACTGATATGAAATTCATATTCATATGTTGACCATCCGCTAAGGCTA	364			

OY 362 CTAGCAGCATATATGAGATCTCCAGCTGCCGCCGATAGTCGAGAACTGATCGGAAT 421
 DB 365 AGGATTTTATCACTTTGTTCTTAAGTTAGGGACCTTAAGATGACCGGGGTCACAC 424
 OY 422 TGTTCAGTATTTAGCCGCGCAAGTACCTCTGTTCCCGGTGCCGGGTGTGATCGGAA 481
 DB 425 TAGCCCCGCTCTTAGCCATTCAGTACACTTTTCCGATCTTGGTATTCATTTGTT 484
 OY 482 TAAGCAGCAGCAGCAGCAGTTCAGATCTCTGTTTGAAGGTTTATGAAGATTGGG 541
 DB 485 TCACATACATCATGTTGTTGGTGAAGGAGACTACTTACAGAGGTTTCATTAAGCGTGG 544
 OY 542 CTTCATCTACTAATTCGAGAGATGATGATCTT 578
 DB 545 CTCTACTCCACAATTCGGTGGACATGACAACTTCT 581
 RESULT 2
 AM221050 583 bp mRNA EST 07-DEC-1999
 LOCUS EST297519 tomato fruit mature green, TAMU Lycopersicon esculentum
 DEFINITION cDNA clone cLEF3F3, mRNA sequence.
 ACCESSION AM221050
 VERSION AM221050.1 GI:6532734
 KEYWORDS EST.
 SOURCE Tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
 I; Solanales; Solanaceae; Solanum; Lycopersicon.
 REFERENCE 1 (bases 1 to 583)
 AUTHORS Alcalá, J., Vrebalov, J., White, R., Mater, A. L., Holt, I. E., Liang, F.,
 Upton, J., Hansen, T., Craven, M. B., Bowman, C. L., Ahn, S., Romling, C. M.,
 Fraser, C. M., Martin, G. B., Tanksley, S. D. and Giovannoni, J.
 Generation of ESTs from tomato fruit tissue
 Unpublished (1999)
 JOURNAL Contact: David Frisch
 COMMENT Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@clemson.edu
 FEATURES
 source
 Location/Qualifiers
 1..583
 /organism="Lycopersicon esculentum"
 /cultivar="T496"
 /db_xref="taxon:4081"
 /clone="cLEF3F3"
 /clone_id="tomato fruit mature green, TAMU"
 /tissue_type="fruit pericarp"
 /dev_stage="mature green (3-5 days pre-ripening)"
 /lab_host="SOLR"
 /note="Vector: pluscript SK(-); Site_1: EcoRI; Site_2:
 XhoI; cLEF - fruit were tagged at the 1cm stage and
 harvested 3-5 days prior to ripening. Fruit were cut in
 half to verify the seeds were indexed. 'Immature' and the
 seeds and locules were discarded prior to freezing the
 pericarp."
 BASE COUNT 142 a 140 c 117 g 184 t
 ORIGIN
 Query Match 9.1%; Score 138.6; DB 20; Length 583;
 Best Local Similarity 52.5%; Pred. No. 7.5e-19;
 Matches 303; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

DB 65 TACTCCCTCTACTTATTTTGGACATGTTTGGTTAGGTTTCCCTATGAGCGGATAT 124
 OY 122 TCTTTAGCACTCCCTGTCCAAACCCGCCCTTCTCCAAACCGTGTCCGAACCTA 181
 DB 125 TATTTTACAAGCTCTCATTTTCAAAACCGGATTTGCTTCAAAACATTTATTCCTCTTA 184
 OY 182 AACAACTTATCTTCTAACCTTCAAAACCTTTTCCCTTTTCAAGCAATCTAATCTACC 241
 DB 185 AAATTCACCTCCCTCCACTCTCAAAACCTATACGCTCTTACCGGAAACGTTGCTTCTC 244
 OY 242 CTCTATCCGCGGAGAAATATGCGGAGTTCGGATGAGACAGTGTGCTTCTTCA 301
 DB 245 CACTAGATACAAAGCATATCTGAGTATGCTATATGACAGAGATTTGTGTGCTTCA 304
 OY 302 CGATTATGAGTCTGTGCGAGATCATCCGATTCGCTCATTAATCTACTGCTTGTCCC 361
 DB 305 CTTTTTTCGAGACTATGATGATTAATTTCAATTTCTCATTTGTTGACCATCCGCTTAAGCGTA 364
 OY 362 CTAGCAGCATATATGAGATCTCCAGCTGCCGCCGATAGTCGAGAACTGATCGGAAT 421
 DB 365 AGGATTTTATCACTTTGTTCTTAAGTTAGGGAACTTAAGATGACCGGGGTCACAC 424
 OY 422 TGTTCAGTATTTAGCCGCGCAAGTACCTCTGTTCCCGGTGCCGGGTGTGATCGGAA 481
 DB 425 TAGCCCCGCTCTTAGCCATTCAGTACACTTTTCCGATCTTGGTATTCATTTGTT 484
 OY 482 TAAGCAGCAGCAGCAGCAGTTCAGATCTCTGTTTGAAGGTTTATGAAGATTGGG 541
 DB 485 TCACATACATCATGTTGTTGGTGAAGACTATACATGATGATTAAGGCGTGGG 544
 OY 542 CTTCATCTACTAATTCGAGAGATGATGATCTT 578
 DB 545 CTCTACTCCACAATTCGGTGGACATGACAACTTCT 581

RESULT 3
 AM102336 687 bp mRNA EST 06-DEC-1999
 LOCUS sdb6d06.y1 Gm-c1009 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-c1009-1284 5' similar to TR:Q9ZWB4 Q9ZWB4 F21M11.13 PROTEIN. ;
 mRNA sequence.
 ACCESSION AM102336
 VERSION AM102336.1 GI:6072949
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Glycine.
 REFERENCE 1 (bases 1 to 687)
 AUTHORS Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna,
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers,
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann,
 R., Waterson, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 JOURNAL Contact: Shoemaker R/Public Soybean EST Project
 COMMENT Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 Worl'd
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 431.

Query Match	7.4%	Score 112	DB 23	Length 518
Best Local Similarity	51.8%	Pred. No. 2.2e-13		
RESULT 5				
AM650280				
LOCUS	518 bp	mRNA	EST	04-APR-2000
DEFINITION	ESR328734 tomato germinating seedlings, TAMU Lycopersicon			
ACCESSION	esculentum cDNA clone cLE112F13 5', mRNA sequence.			
VERSION	AM650280			
KEYWORDS	AM650280.1 GI:7411518			
SOURCE	EST.			
ORGANISM	tomato.			
	Lycopersicon esculentum			
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
	Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids			
	I; Solanales; Solanaceae; Solanum; Lycopersicon.			
REFERENCE	1 (bases 1 to 518)			
AUTHORS	Alcalai, J., Veshalov, J., White, R., van der Hoeven, R.S., Holt, I.E., Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Ronning, C.M., Nierman, W., Fraser, C.M., Giovannoni, J.D., Martin, G.B. and Tanksley , S.D.			
TITLE	Generation of ESTs from germinating tomato seed			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: David Frisch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU			
FEATURES	5 prime sequence.			
source	Location/Qualifiers			
	1..518			
	/organism="Lycopersicon esculentum"			
	/cultivar="TA496"			
	/db_xref="taxon:4081"			
	/clone="cLE112F13"			
	/clone_1fb="tomato germinating seedlings, TAMU"			
	/tissue_type="whole seedlings"			
	/dev_stage="7 days post imbibition"			
	/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:			
	XhoI; 7 days post imbibition on water-jagar. Mixed stage			
	whole germinating seedlings from seed coat emergence up			
	to two centimeters in seeds not showing obvious signs of			
	germination were discarded."			
BASE COUNT	134 a	119 c	100 g	165 t
ORIGIN				

[illegible]

TITLE Generation of ESTs from tomato leaf tissue
JOURNAL Unpublished (1999)
COMMENT Contact: David Fritsch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@clemson.edu

FEATURES
source Location/Qualifiers
1..490
/organism="Lycopersicon esculentum"
/cultivar="Rio Grande P10R"
/db_xref="taxon:4081"
/clone="cLET6D22"
/clone_11b="tomato mixed elicitor, BRT"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI, Site_2: XhoI; cLET - Inoculated with a variety of disease response elicitors. Plants exposed to 2,6 dichloroisonicotinic acid, BTH, jasmonic acid, ethylene, fenchion, EIX, okadaic acid, or systemin prior to tissue harvest. EcoRI site was destroyed during cloning."

BASE COUNT 137 a 103 c 93 g 157 t
ORIGIN

Query Match 5.6%; Score 84.8; D3 19; Length 490;
Best Local Similarity 55.4%; Pred. No. 8.9e-08;
Matches 164; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

```

Oy 28 GTGGGCGCCGCTCCAGGCGGTGCTGAGCTACCTCCGCTACCTTTGGACATG 87
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 GGGCCACCTCCCGAGGGCGGCGCATGCTACGCTCCCTCTAATTTGATCAT 63

Oy 88 AGTGGCTGCATTTCCACCCCATGCTTCTCTCTTACAGACTCCCTGTTCCAAA 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 ATGTGTTGTTTGGCTATATGCTAGATTTATTTCTAACCTATTCGATTTCCAAA 123

Oy 148 CCGGCTTCCTCGAAGCCGTGCTCCGAAACTCAATCTTATCTTAACCTCCAAA 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 CTCGATTTGTTCAAAACATTTATCTTCTTAAACATCTCTCCCTCATCTCAAA 183

Oy 208 CACTTCTTCCCTTTCATGCAATCTAATCTACCTCTCTGCGGAGAAATGCGGAG 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 CACTACACACCTTAGTTGGCAACATTTGCATGCTCCACTAATTTCTAGTGTATTCAGAG 243

Oy 268 TTCCGGTATCAGAACGGTGAATCGTTCTTTCACGATTTATGAGTGTGCGAGA 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 CTGTGTATGTGACTGAGATTTCTGTATCAGATTTCTTACTGAACATGACATGGA 299

```

RESULT 9
AM093479 409 bp mRNA EST 18-OCT-1999
LOCUS EST286659 tomato mixed elicitor, BRT Lycopersicon esculentum CDNA
DEFINITION clone cLET25A3, mRNA sequence.
ACCESSION AM093479
VERSION AM093479.1 GI:6059074
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 409)
AUTHORS D'Ascenzo, M., He, X., Lyman, J., Holt, J. E., Liang, F., Upton, J.,
Roaming, C. M., Craven, M. B., Fujii, C. Y., Bowman, C. L., Niernman, W.,
Fraser, C. M., Venter, J. C., Martin, G. B., Tanksley, S. D. and Giovannoni,
J.

TITLE Generation of ESTs from tomato leaf tissue
JOURNAL Unpublished (1999)
COMMENT Contact: David Fritsch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@clemson.edu

FEATURES
source Location/Qualifiers
1..409
/organism="Lycopersicon esculentum"
/cultivar="Rio Grande P10R"
/db_xref="taxon:4081"
/clone="cLET25A3"
/clone_11b="tomato mixed elicitor, BRT"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI, Site_2: XhoI; cLET - Inoculated with a variety of disease response elicitors. Plants exposed to 2,6 dichloroisonicotinic acid, BTH, jasmonic acid, ethylene, fenchion, EIX, okadaic acid, or systemin prior to tissue harvest. EcoRI site was destroyed during cloning."

BASE COUNT 108 a 91 c 74 g 136 t
ORIGIN

Query Match 5.5%; Score 83.2; DB 19; Length 409;
Best Local Similarity 55.1%; Pred. No. 1.9e-07;
Matches 163; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

```

Oy 28 GTGGGCGCCGCTCCAGGCGGTGCTGAGCTACCTCCGCTACCTTTGGACATG 87
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14 GGGCCACCTCCCGAGGGCGGCGCATGCTACGCTCCCTCTAATTTGATCAT 73

Oy 88 AGTGGCTGCATTTCCACCCCATGCTTCTCTCTTACAGACTCCCTGTTCCAAA 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74 ATGTGTTGTTTGGCTATATGCTAGATTTATTTCTAACCTATTCGATTTCCAAA 133

Oy 148 CCGGCTTCCTCGAAGCCGTGCTCCGAAACTCAATCTTATCTTAACCTCCAAA 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 CTCGATTTGTTCAAAACATTTATCTTCTTAAACATCTCTCCCTCATCTCAAA 193

Oy 208 CACTTCTTCCCTTTCATGCAATCTAATCTACCTCTCTGCGGAGAAATGCGGAG 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 CACTACACACCTTAGTTGGCAACATTTGCATGCTCCACTAATTTCTAGTGTATTCAGAG 253

Oy 268 TTCCGGTATCAGAACGGTGAATCGTTCTTTCACGATTTATGAGTGTGCGAGA 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 CTGTGTATGTGACTGAGATTTCTGTATCAGATTTCTTACTGAACATGACATGGA 309

```

RESULT 10
BE187619 788 bp mRNA EST 22-JUN-2000
LOCUS EST336180 KY0 Medicago truncatula cDNA clone PKY0-1613, mRNA
DEFINITION sequence.
ACCESSION BE187619
VERSION BE187619.1 GI:8666803
KEYWORDS EST.
SOURCE barbel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosoids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 788)
AUTHORS VandenBosch, K., Hur, J., Beremand, P., Peng, H. and Ellis, L.
TITILE ESTs from uninoculated seedling roots of Medicago truncatula
JOURNAL Unpublished (1999)
COMMENT Contact: VandenBosch K

56 AGCAGTCACTCCCGCTCACCTTCTTGACATGACGTGGCTGCATTTCACCCCATGCTTC 115
| | | | | | | | | | | | | | | | | | | | |
81 AACTATCTCTCCCTCTGCACATTTTGTGACATATATTGGCTAAGATTTTCATCCGGTAGAAC 140

```

/organism="Medicago truncatula/Glomsus versiforme mixed EST
library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="PMHAM-56M23"
/clone_1b="MHAM"
/tissue_type="roots colonized with Glomsus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days

```


JOURNAL DNA Res. 7, 175-180 (2000)
COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers

FEATURES
source

1. 605
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="Ap217a04_f"
/clone_lib="Arabidopsis thaliana above-ground organ two to
six-week old"
/tissue_type="above-ground organ"
/dev_stage="two to six-week old"
BASE COUNT 202 a 168 c 86 g 149 t
ORIGIN

Query Match 4.8%; Score 73.4; DB 18; Length 605;

Best Local Similarity 49.0%; Pred. No. 2e-05;
Matches 225; Conservative 0; Mismatches 231; Indels 3; Gaps 1;

```
OY 908 TCGATTTAAGCCCGCTTGGATCCGCCGCGCTCCGGGAACCTTCGGAACCTGCTAT 967
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 498 TGGATTGTCGAAGCCTTATGCTCCACCGGTTCCATCGAGTATTGTTGTTATTTGTTAT 439

OY 968 CATTTCGATGGCGAAGATCCTGCGCGGATTTGGTCGAGATGAAGGGGTGTTCCGG 1027
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 438 CTGCTTTGTTTAAATGTCGTTAACGCCAGACATTTTATGATGTAGAGAGAGGTTTATG 379

OY 1028 CAGCTGAGCCGATCCGCCGCGAATAAGAGAGACGACACAGAAGATTCTAGAAA 1087
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 378 CTGCTGCGAAGATGGTTAGTATCTGTTGAGGCATTTGATGAGATGTTGCATTGAAGA 319

OY 1088 CTGTGAGAACTGCGCGTCTGAGATTTCGGAAGCCTTCGAAACTGTTATTCTCGGTGG 1147
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 318 TTCCAGAGATTTTGAAGGTTTACGACTCTTTCACAGAACACAGATTGTTCTGTTG 259

OY 1148 CGGATTCGAGCAGGCTTGATCTTTACGGCGCGATTTTGGATGGGTAAGCGGTGAAGC 1207
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 258 CCGGCTCGACCGGTTCCGAGTTTACGGTTGATTTTGGGTGGGTAGACCGGAGAAAG 199

OY 1208 AAGAGATCTGTCGATTTATGAGAGAACTTACGATGTCGTTGTAAACCGAGGATG 1267
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 198 TGGTGGTGTGTCATTGACCAAGGTGA--AGCATTTCTTTGCGGAGGTAGAGATG 142

OY 1268 CTGCCGAGAGATTGAGGTTGATGTTGCCAAAGAGAGATTGCAAGCTTTTGATG 1327
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 141 GGAGTGTGTGTAGAGCTTGCTTCCCTCAAGAAACATGAATGATGTTCTTGTG 82

OY 1328 ATTATTTTGCAGAGGAATAAGGGTGAATTAATCATTT 1366
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 81 ATTTCCTTCACAAAGGACTAGAAATTAATCATTTTTT 43
```

Search completed: October 20, 2000, 02:07:32
Job time: 1915 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2000, 01:34:52 ; Search time 16.58 Seconds
(without alignments)
33.790 Million cell updates/sec

Title: US-08-894-356c-21
Perfect score: 40
Sequence: 1 DFGWCK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues
Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_14:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	331	10	004457 arbidops
2	40	100.0	436	10	004458 arbidops
3	40	100.0	439	10	092TK5 catharant
4	40	100.0	443	10	09SMD9 arbidops
5	40	100.0	450	10	09SMD9 arbidops
6	40	100.0	451	10	064470 arbidops
7	40	100.0	455	10	P93094 cucumis mel
8	40	100.0	460	10	043583 nicotiana t
9	40	100.0	461	10	048761 arbidops
10	40	100.0	469	10	092WR8 gentiana tr
11	37	92.5	99	10	023943 fragaria ve
12	37	92.5	431	10	09SST8 ipomoea bat
13	37	92.5	482	10	004201 arbidops
14	36	90.0	280	10	081365 prunus arme
15	36	90.0	430	10	09SMW7 arbidops
16	35	87.5	430	10	09SMD2 clarkia con
17	35	87.5	431	10	09SMD1 clarkia con
18	35	87.5	431	10	09SMD0 clarkia con
19	35	87.5	433	10	064988 clarkia bre

20	35	87.5	433	10	09SPU3 clarkia bre
21	35	87.5	442	10	023916 dianthus ca
22	35	87.5	445	10	024645 dianthus ca
23	35	87.5	445	10	023918 dianthus ca
24	35	87.5	446	10	023917 dianthus ca
25	35	87.5	451	10	09SHQ2 arbidops
26	35	87.5	476	10	09SL56 arbidops
27	35	87.5	485	5	09TYR1 caenorhabd
28	35	87.5	572	10	064549 arbidops
29	34	85.0	113	10	042482 populus del
30	34	85.0	216	10	09SUC8 arbidops
31	34	85.0	361	4	075654 homo sapien
32	34	85.0	361	4	09YRT5 homo sapien
33	34	85.0	493	10	09SKS9 arbidops
34	34	85.0	741	10	022143 arbidops
35	33	82.5	137	10	P93856 skeletonema
36	33	82.5	190	10	P93855 skeletonema
37	33	82.5	730	2	09RWC0 delinococcus
38	32	80.0	234	2	09ZEP4 streptomyce
39	32	80.0	271	9	09T172 bacterioph
40	32	80.0	331	2	084959 raietonia s
41	32	80.0	331	2	09RAF8 raietonia s
42	32	80.0	346	10	024328 ricinus com
43	32	80.0	367	10	043445 glycine max
44	32	80.0	409	5	093323 caenorhabd
45	32	80.0	595	5	09W4K2 drosophila

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	331 AA.
004457	004457			
AC	004457			
DT	01-JUL-1997 (TREMBlrel. 04, Created)			
DT	01-JUL-1997 (TREMBlrel. 04, Last sequence update)			
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)			
DE	F21J9.19.			
OS	Arbidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;			
OC	Brassicaceae; Arbidopsis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Dewar K., Buehler E., Feng J., Kim C., Li Y., Shinn P., Sun H.,			
RA	Conway A., Conway A., Kurtz D., Oji O., Osborne B., Shen Y. K.,			
RA	Torlund M., Vysotskaya V., Yu G., Davis R.W., Federspiel N.A.,			
RA	Theologis A., Ecker J.R.,			
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC000103; AAB61522.1;			
DR	MEDDEL; 16800; Arath; 2598; 16800.			
SQ	SEQUENCE 331 AA; 37598 MW; 9E26B7F5D1DD1410 CRC64;			
Qy	1 DFGWCK 6			
Db	265 DFGWCK 270			
Query Match	100.0%;	Score 40;	DB 10;	Length 331;
Best Local Similarity	100.0%;	Pred. No. 9.6;		
Matches	6;	Conservative 0;	Mismatches 0;	Indels 0;
RESULTS	2			
ID	004458	PRELIMINARY:	PRT:	436 AA.
AC	004458			
DT	01-JUL-1997 (TREMBlrel. 04, Created)			
DT	01-JUL-1997 (TREMBlrel. 04, Last sequence update)			
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)			
DE	F21J9.20.			
OS	Arbidopsis thaliana (Mouse-ear cress).			

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA Dewar K., Buehler E., Peng J., Kim C., Li Y., Shin P., Sun H.,
RA Conway A., Kurtz D., Oji O., Osborne B., Shen Y.K.,
RA Toriumi M., Vysotskaya V., Yu G., Davis R.W., Federspiel N.A.,
RA Theologis A., Ecker J.R.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC000103; AAB61523.1; -
DR MENDEL: 16799; Arah1;2597;16799.
SQ SEQUENCE 436 AA; 48824 MW; B303A52DE3FD1355 CRC64;

Query Match 100.0%; Score 40; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGR 6
|||||
DB 372 DFGWGR 377

RESULT 3
ID Q9ZTK5 PRELIMINARY; PRT; 439 AA.
AC Q9ZTK5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE DEACETYLINDOLINE 4-O-ACETYLTRANSFERASE (EC 2.3.1.107).
GN DAT.
OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Gentianales;
OC Apocynaceae; Catharanthus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. LITTLE DELICATA;
RX MEDLINE: 98346012.
RA St-Pierre B., Laflamme P., Alarco A.M., De Luca V.;
RT "The terminal O-acetyltransferase involved in vindoline biosynthesis
RT defines a new class of proteins responsible for coenzyme A-dependent
RT acyl transfer.";
RL Plant J. 14:703-713(1998).
DR EMBL: AF053307; AAC99311.1; -
KW Transferase; Acyltransferase.
SQ SEQUENCE 439 AA; 49873 MW; 1357C76550E0D04D CRC64;

Query Match 100.0%; Score 40; DB 10; Length 439;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGR 6
|||||
DB 380 DFGWGR 385

RESULT 4
ID Q9SND9 PRELIMINARY; PRT; 443 AA.
AC Q9SND9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ANTHRANILATE N-HYDROXYCINNAMOYL/BENZOYLTRANSFERASE-LIKE PROTEIN.
GN F11C1.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsis.

RN [1]
RP SEQUENCE FROM N.A.
RA Barques M., Collado M.C., Navarro P., Terol J., Perez-Alonso M.,
RA Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132976; CAB62307.1; -
KW Transferase.
SQ SEQUENCE 443 AA; 49960 MW; C4ADB67A0E3AB3D4 CRC64;

Query Match 100.0%; Score 40; DB 10; Length 443;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGR 6
|||||
DB 386 DFGWGR 391

RESULT 5
ID Q9SNE0 PRELIMINARY; PRT; 450 AA.
AC Q9SNE0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ANTHRANILATE N-HYDROXYCINNAMOYL/BENZOYLTRANSFERASE-LIKE PROTEIN.
GN F11C1.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA Barques M., Collado M.C., Navarro P., Terol J., Perez-Alonso M.,
RA Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132976; CAB62306.1; -
KW Transferase.
SQ SEQUENCE 450 AA; 49805 MW; 5D4A4AAABCD3F0C6 CRC64;

Query Match 100.0%; Score 40; DB 10; Length 450;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGR 6
|||||
DB 391 DFGWGR 396

RESULT 6
ID O64470 PRELIMINARY; PRT; 451 AA.
AC O64470;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE PUTATIVE ANTHRANILATE N-HYDROXYCINNAMOYL/BENZOYLTRANSFERASE.
GN T20R24.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsis.
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,
 RA Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,
 RA Somerville C.R., Venter J.C.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002392; AAD12025.1; -
 KW Transferrase.
 RN
 SQ SEQUENCE 451 AA; 50434 MW; B13B2CBF4D4B1482 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 40; DB 10; Length 451;
 Pred. No. 13;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
 Db 397 DFGWGK 402

RESULT 7
 ID P93094 PRELIMINARY; PRT; 455 AA.
 AC P93094;
 DT 01-MAY-1997 (TREMUREL. 03, Created)
 DT 01-MAY-1997 (TREMUREL. 03, Last sequence update)
 DT 01-NOV-1999 (TREMUREL. 12, Last annotation update)
 DE HYPOHETICAL 50.8 KDA PROTEIN (FRAGMENT).
 OS Cucumis melo (Muskmelon).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Cucurbitales;
 OC Cucurbitaceae; Cucumis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CANTALOUPE CHARENTAIS; TISSUE-FRUIT;
 RX MEDLINE: 97188564.
 RA Aggelis A., John I., Karvouni Z., Grierson D.;
 RT Characterization of two cDNA clones for mRNAs expressed during
 RT ripening of melon (Cucumis melo L.) fruits.;
 RL Plant Mol. Biol. 33:313-322(1997).
 DR EMBL: 270521; CAA94432.1; -
 DR MENDEL: 16795; Cucme; 2594.16795.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 455 AA; 50775 MW; EAC68CAE7459F924 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 40; DB 10; Length 455;
 Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
 Db 374 DFGWGK 379

RESULT 8
 ID Q43583 PRELIMINARY; PRT; 460 AA.
 AC Q43583;
 DT 01-NOV-1996 (TREMUREL. 01, Created)
 DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)
 DT 01-NOV-1999 (TREMUREL. 12, Last annotation update)
 DE HSR201 PROTEIN.
 GN HSR201.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
 OC Solanaceae; Nicotiana.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. BOTTOM SPECIAL; TISSUE-LEAF;
 RX MEDLINE: 96343929.

RA Czerwiec P., Huang H.C., Marco Y.;
 RT "Characterization of hsr201 and hsr515, two tobacco genes
 RT preferentially expressed during the hypersensitive reaction provoked
 RT by phytopathogenic bacteria.";
 RL Plant Mol. Biol. 31:255-265(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. BOTTOM SPECIAL; TISSUE-LEAF;
 RA Yves M.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X95343; CAA64636.1; -
 DR MENDEL: 16796; Nicra; 2594.16796.
 SQ SEQUENCE 460 AA; 50972 MW; F67456B8C6F0E3D1 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 40; DB 10; Length 460;
 Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
 Db 382 DFGWGK 387

RESULT 9
 ID 048761 PRELIMINARY; PRT; 461 AA.
 AC 048761;
 DT 01-JUN-1998 (TREMUREL. 06, Created)
 DT 01-JUN-1998 (TREMUREL. 06, Last sequence update)
 DT 01-AUG-1998 (TREMUREL. 07, Last annotation update)
 DE F21B7.32.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
 OC Brassicaceae; Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Buehler E., Dewar K., Feng J., Kim C., Li Y., Sun H.,
 RA Conway A., Conway A., Kurtz D., Oji O., Shen Y.K., Toriumi M.,
 RA Vysotskaia V., Yu G., Davis R.W., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002560; AAB97723.1; -
 SQ SEQUENCE 461 AA; 52086 MW; DA6CB03CD40B82F4 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 40; DB 10; Length 461;
 Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
 Db 400 DFGWGK 405

RESULT 10
 ID Q92WR8 PRELIMINARY; PRT; 469 AA.
 AC Q92WR8;
 DT 01-MAY-1999 (TREMUREL. 10, Created)
 DT 01-MAY-1999 (TREMUREL. 10, Last sequence update)
 DT 01-JUN-2000 (TREMUREL. 14, Last annotation update)
 DE ANTHOCYANIN 5-AROMATIC ACTYLTRANSFERASE.
 OS Gentiana triflora.
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Gentianales;
 OC Gentianaceae; Gentiana.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PETAL;
 RX MEDLINE: 99097837.
 RA Fujiwara H., Tanaka Y., Yonekura-Sakakibara K., Fukuchi-Mizutani M.,

RA Nakao M., Fukui Y., Yamaguchi M., Ashikari T., Kusumi T.;
 RT "cDNA cloning, gene expression and subcellular localization of
 RT anthocyanin 5-aromatic acyltransferase from Gentiana triflora.";
 RL Plant J. 16:421-431(1998).
 DR EMBL; AB010708; CAB19224.1; -.
 KW Transferase; Acyltransferase.
 SQ SEQUENCE 469 AA; 52739 MW; EDEFFB0B5E5963ACE CRC64;

Query Match 100.0%; Score 40; DB 10; Length 469;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DFGWGR 6
 DB 411 DFGWGR 416

RESULT 11
 ID 023943 PRELIMINARY; PRT; 99 AA.
 AC 023943;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
 DE RIPENING-INDUCED PROTEIN (FRAGMENT).
 OS Fragaria vesca.
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Rosales; Rosaceae;
 OC Fragaria.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. REINE DES VALLEES;
 RA Nam Y.W., Tichit L., Lepeletier M., Guerg B., Marty I., Lelievre J.M.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A001450; CAA04771.1; -.
 FT NON_TER 1
 SQ SEQUENCE 99 AA; 11336 MW; 3B567A4617095858 CRC64;

Query Match 92.5%; Score 37; DB 10; Length 99;
 Best Local Similarity 83.3%; Pred. No. 8.6;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DFGWGR 6
 DB 32 DFGWGR 37

RESULT 12
 ID 09SST8 PRELIMINARY; PRT; 431 AA.
 AC 09SST8;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE N-HYDROXYCINNAMOYL/BENZOYLTRANSFERASE.
 GN HCBT.
 OS Ipomoea batatas (Sweet potato) (Batace).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
 OC Convolvulaceae; Ipomoea.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kikuchi Y., Shiohara K., Morita H., Kashiwagi K., Hsu T., Noguchi H.;
 RT "Ipomoea batatas hcbt mRNA for N-hydroxycinnamoyl/benzoyltransferase
 RT synthase, complete cds."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB035183; BAA87043.1; -.
 KW Transferase.
 SQ SEQUENCE 431 AA; 47929 MW; E161971850B68C32 CRC64;

Query Match 92.5%; Score 37; DB 10; Length 431;
 Best Local Similarity 83.3%; Pred. No. 43;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DFGWGR 6
 DB 378 DFGWGR 383

RESULT 13
 ID 004201 PRELIMINARY; PRT; 482 AA.
 AC 004201;
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE HYPOTHETICAL 52.9 KDA PROTEIN T29M21.14 IN CHROMOSOME II.
 GN T29M21.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
 OC Brassicaceae; Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Rounsley S.D., Ketchum K.A., Ian X., Phillips C.A., Brandon R.C.,
 RA Fuhrmann J.L., White O., Kerlavage A.R., Adams M.D., Somerville C.R.,
 RA Venter J.C., (1997) to the EMBL/GenBank/DBJ databases.
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF002109; AAB95283.1; -.
 DR MENDEL; 16798; Arath; 2596; 16798.
 KW Hypothetical protein.
 SQ SEQUENCE 482 AA; 52882 MW; 3C6341771A81989A CRC64;

Query Match 92.5%; Score 37; DB 10; Length 482;
 Best Local Similarity 83.3%; Pred. No. 48;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DFGWGR 6
 DB 422 DFGWGR 427

RESULT 14
 ID 081365 PRELIMINARY; PRT; 280 AA.
 AC 081365;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE AP2 DOMAIN CONTAINING PROTEIN (FRAGMENT).
 GN AP2DCP.
 OS Prunus armeniaca (Apriocot).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Rosales; Rosaceae;
 OC Prunus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERGERON; TISSUE-MESOCARP, EXOCARP;
 RA Mbegule-A-Mbegule D., Fils-Lycaon B.;
 RT "Molecular cloning and partial nucleotide sequence of an AP2 domain
 RT containing protein from apriocot."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF071893; AAC24587.1; -.
 DR INTERPRO; IPR001471; -.
 DR PFAM; PF00847; Ap2-domain; 1.
 FT NON_TER 1
 SQ SEQUENCE 280 AA; 31161 MW; 5E6289798F846D44 CRC64;

Query Match 90.0%; Score 36; DB 10; Length 280;
 Best Local Similarity 83.3%; Pred. No. 40;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
|||||:

Db 169 DFGWGE 174

RESULT 15

09SMW7

PRELIMINARY; PRT; 430 AA.

AC 09SMW7;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE HYPOTHETICAL 48.0 KDA PROTEIN.

GN T8P19.230.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;

OC Brassicaceae; Arabidopsis.

RN [1]

RP SEQUENCE FROM N.A.

RA Choisme N., Robert C., Brottier P., Wincker P., Catolico L.,

RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,

RA Mayer K.F.X., Queller F., Salanoubat M.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Arabidopsis sequencing project;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL133315; CAB62361.1; -

KM Hypothetical protein.

SO SEQUENCE 430 AA; 48004 MW; 6210941B7C14B31 CRC64;

Query Match 90.0%; Score 36; DB 10; Length 430;
Best Local Similarity 83.3%; Pred. No. 64;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFGWGK 6
|||||:

Db 378 DFGWGE 383

Search completed: October 20, 2000, 01:36:48
Job time: 116 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2000, 03:37:56 ; Search time 3710.09 seconds

(without alignments)
20.013 Million cell updates/sec

Title: US-08-894-356c-22

Perfect score: 17

Sequence: 1 GAYTTGGNTGGGNA 17

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sy:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pr4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pr5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_v11:*
73: gb_v12:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pr6:*
92: gb_pr7:*
93: gb_sta1:*
94: gb_sta2:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14.2	83.5	850	7	AF227981
2	14.2	83.5	1080	5	AF16155
3	14.2	83.5	1080	5	A20628
4	14.2	83.5	1080	5	AR003699
5	14.2	83.5	1080	5	I28271
6	14.2	83.5	1320	7	AF053307
7	14.2	83.5	1419	7	AF190130
8	14.2	83.5	1476	7	AB029340
9	14.2	83.5	1479	5	E12756
10	14.2	83.5	1508	5	E12757
11	14.2	83.5	1512	8	CPMPEL2
12	14.2	83.5	1526	5	A65937

13 14.2 83.5 1542 5 116758 116758 Sequence 1
14 14.2 83.5 1563 48 NTHSR201 X95343 N.tabacum m
15 14.2 83.5 1605 5 E12755 Petunia hybrida
16 14.2 83.5 1605 7 AB026495 Petunia x
17 14.2 83.5 1622 5 E12754 Gentiana
18 14.2 83.5 1622 7 AB026494 Gentiana
19 14.2 83.5 1679 7 AB010708 Gentiana
20 14.2 83.5 1703 5 E12753 Gentiana
21 14.2 83.5 1735 8 DCHCWR3 284571 D.caryophyll
22 14.2 83.5 3706 5 I16766 Sequence 16
23 14.2 83.5 4746 4 AF228714 I16766 Sequence 16
24 14.2 83.5 8222 2 LLA250129 AF228714 Ictulurus
25 14.2 83.5 10238 74 U32728 AF228714 Ictulurus
26 14.2 83.5 12850 1 AF282249 U32728 Haemophilus
27 14.2 83.5 16078 43 AC019969 AF282249 Lactococcus
28 14.2 83.5 17500 42 AC011779 AC019969 Drosophila
29 14.2 83.5 59688 40 AC011588 AC011779 Drosophila
30 14.2 83.5 66237 7 AB016892 AC011588 Homo sapi
31 14.2 83.5 74342 7 AB020742 AB016892 Arabidops
32 14.2 83.5 74342 43 AC020525 AC020742 Arabidops
33 14.2 83.5 78174 32 AC004280 AC020525 Arabidops
34 14.2 83.5 79676 7 AB013396 AB004280 Drosophila
35 14.2 83.5 82360 8 AP000606 AB013396 Arabidops
36 14.2 83.5 83650 7 AB023041 AP000606 Arabidops
37 14.2 83.5 86652 44 AC005131 AB023041 Arabidops
38 14.2 83.5 87581 8 AT77H20 AL162508 Arabidops
39 14.2 83.5 87841 7 AB005247 AB005247 Arabidops
40 14.2 83.5 88318 8 AP002842 AP005247 Oryza sat
41 14.2 83.5 95477 10 AC007076 AC002842 Oryza sat
42 14.2 83.5 98259 7 AC002392 AC007076 Homo sapi
43 14.2 83.5 98771 53 AC025044 AC002392 Arabidops
44 14.2 83.5 99392 7 AC000103 AC025044 Oryza sat
45 14.2 83.5 99392 7 AC000103 AC000103 Genomic s

ALIGNMENTS

RESULT 1
LOCUS AF227981 850 bp mRNA PLN 17-FEB-2000
DEFINITION Euphorbia esula F21J9.20-like protein mRNA, partial cds.
ACCESSION AF227981
VERSION AF227981.1 GI:6984225
KEYWORDS
SOURCE
ORGANISM . leafy spurge.
Euphorbia esula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Malpighiales; Euphorbiaceae; Euphorbia.
REFERENCE
AUTHORS 1 (bases 1 to 850)
TITLE Anderson, J.V. and Horvath, D.P.
of Euphorbia esula (leafy spurge)
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 850)
TITLE Anderson, J.V. and Horvath, D.P.
Direct Submission
JOURNAL Submitted (26-JAN-2000) Plant Science, USDA/ARS, 1605 Albrecht
Blvd., Fargo, ND 58105, USA
FEATURES
source
1..850
/organism="Euphorbia esula"
/db_xref="taxon:3993"
/clone="12a"
/tissue_type="underground adventitious buds"
/dev_stage="3-day induced (defoliated)"
CDS
1..660
/note="similar to Arabidopsis thaliana F21J9.20 encoded by
GenBank Accession Number AC000103"
/codon_start=1
/evidence="not_experimental"
/product="F21J9.20-like protein"

/protein_id="AAF34801.1"
/db_xref="GI:6984225"
/translation="RRFVDSKIGALKEIASAIAAPKPTREAVSGLIMKLTIAFK
KSNPUIRSTVSASVNNLRPTPPVPPENHAGNLVITPKVDEAMELKGIVIKOG
MODFENVYKRVQEDVGALICEFGRFAEKLSKIDFPGSGMCRGLVDADFGWG
KPTWLSIVSTINRVCIILDTKDGGFEMWTLSEBDSMFESDERVLEFAQVNGVT
L"
3'UTR 661..850
polya_site 827
BASE COUNT 258 a 123 c 227 g 242 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 7; Length 850;
Best Local Similarity 76.5%; Pred. No. 7.9e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAYTTYGNTGGGNA 17
DB 466 GATTTCGGTGGGNA 482

RESULT 2
LOCUS A16155 1080 bp DNA PAT 03-OCT-1994
DEFINITION PTOM36.
ACCESSION A16155
VERSION A16155.1 GI:640912
KEYWORDS
SOURCE . synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1080)
ARTIFICIAL SEQUENCE.

AUTHORS
TITLE TRANSGENIC PLANTS WITH INCREASED SOLIDS CONTENT
JOURNAL Patent: WO 9314212-A 1 22-JUL-1993;
FEATURES
source
1..1080
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 356 a 141 c 208 g 375 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 5; Length 1080;
Best Local Similarity 76.5%; Pred. No. 7.8e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAYTTYGNTGGGNA 17
DB 866 GATTTCGATGGGNA 882

RESULT 3
LOCUS A20628 1080 bp DNA PAT 12-JUN-1994
DEFINITION PTOM36 fragment.
ACCESSION A20628
VERSION A20628.1 GI:583390
KEYWORDS
SOURCE . synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1080)
ARTIFICIAL SEQUENCE.

AUTHORS
TITLE DNA, CONSTRUCTS, CELLS AND PLANTS DERIVED THEREFROM
JOURNAL Patent: WO 9105865-A 1 02-MAY-1991;
FEATURES
source
1..1080
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 356 a 141 c 208 g 375 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 5; Length 1080;
Best Local Similarity 76.5%; Pred. No. 7.8e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGGNTGGGNA 17
||:|||||
Db 866 GATTTTGGATGGGNA 882

RESULT 4
LOCUS AR003699 1080 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5744364.
ACCESSION AR003699
VERSION AR003699.1 GI:3964958
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1080)
AUTHORS Bird,C.R., Glieron,D., Ray,J.,Anthony and Schuch,W.Walter.
TITLE PROM36 constructs and tomato cells transformed therewith
JOURNAL Patent: US 5744364-A 1 28-APR-1998;
FEATURES
source location/Qualifiers
1..1080
/organism="unknown"
BASE COUNT 356 a 141 c 208 g 375 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 5; Length 1080;
Best Local Similarity 76.5%; Pred. No. 7.8e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGGNTGGGNA 17
||:|||||
Db 866 GATTTTGGATGGGNA 882

RESULT 5
LOCUS 128271 1080 bp DNA PAT 06-FEB-1997
DEFINITION Sequence 1 from patent US 5569829.
ACCESSION 128271
VERSION 128271.1 GI:1819047
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1080)
AUTHORS Bird,C.R., Bonnell,J.M., Glieron,D., Ray,J.A. and Schuch,W.W.
TITLE Transformed tomato plants
JOURNAL Patent: US 5569829-A 1 29-OCT-1996;
FEATURES
source location/Qualifiers
1..1080
/organism="unknown"
BASE COUNT 356 a 141 c 208 g 375 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 5; Length 1080;
Best Local Similarity 76.5%; Pred. No. 7.8e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGGNTGGGNA 17
||:|||||
Db 866 GATTTTGGATGGGNA 882

RESULT 6
AF053307

LOCUS AF053307 1320 bp DNA PLN 01-JAN-1999
DEFINITION Catharanthus roseus deacetylindoline 4-O-acetyltransferase (DAT)
ACCESSION AF053307
VERSION AF053307.1 GI:4091807
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1320)
AUTHORS St-Pierre,B., Laflamme,P., Alarco,A.M. and De Luca,V.
TITLE The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer
JOURNAL Plant J. 14 (6), 703-713 (1998)
MEDLINE 98346012
AUTHORS St-Pierre,B., Laflamme,P. and De Luca,V.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1998) Institut de Recherche en Biologie Vegetale, 4101 Sherbrooke East, Montreal, QC H1X 2B2, Canada
FEATURES
source location/Qualifiers
1..1320
/organism="Catharanthus roseus"
/culturvar="Little Delicata"
/db_xref="taxon:4058"
<1..>1320
/gene="DAT"
/product="deacetylindoline 4-O-acetyltransferase"
<1..>1320
/gene="DAT"
1..1320
/gene="DAT"
/EC_number="2.3.1.107"
/function="involved in the last step in vindoline biosynthesis"
/note="acetylcoenzyme A:deacetylindoline 17-O-acetyltransferase"
/codon_start=1
/product="deacetylindoline 4-O-acetyltransferase"
/protein_id="AAC9311.1"
/db_xref="GI:4091808"
/translation="MESGKISVETETSKYLIRKSSPTQSLSRVNLVNDQNTVQTC VSGPFYENPDGIEISTIREQLQNSLSTLVSYVPFAGKVKNDYIHGNDGIEFVEV RIRCRMDILKYELEISYARDLVLPKRVYVSGEDTATVQLSHPCGLAVAFGISHEV ADGSTIASFMKDMAASACVLSSSHHPVPLPLVSDSIIPRODNIICEOPTKNCVKEK FIPPEPAIERKSKRAVETGTEKPTREVEPLTFLSRCAVACKSAKNNKCGSLPPVY LQAINLRPIELPONSVGNLVYSIFSRIRIKENDVLINEKETVRLVINELRKOKIKNL SREKLVTVQMEERFVSKLEFDISNFDIDVLYSDSWCRPFYVDVDFWGRPIWVCLF QPYIKNCVVMNDYPPGDYGEIAIVSFQEKMSAFKNEQLLOFVSN"
BASE COUNT 409 a 244 c 252 g 415 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 7; Length 1320;
Best Local Similarity 76.5%; Pred. No. 7.7e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGGNTGGGNA 17
||:|||||
Db 1138 GATTTTGGATGGGNA 1154

RESULT 7
AF190130 1419 bp mRNA PLN 15-FEB-2000
LOCUS AF190130
DEFINITION Taxus cuspidata taxadienol acetyl transferase (TAT) mRNA, complete cds.
ACCESSION AF190130
VERSION AF190130.1 GI:6978037
KEYWORDS

SOURCE Taxus cuspidata.
ORGANISM Taxus cuspidata
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Coniferales; Taxaceae; Taxus.

REFERENCE 1 (bases 1 to 1419)
AUTHORS Walker, K., Schoendorf, A. and Croteau, R.
TITLE Molecular Cloning of a Taxa-4(20),11(12)-dien-5(alpha)-ol-O-Acetyl transferase cDNA from Taxus and Functional Expression in Escherichia coli
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1419)
AUTHORS Walker, K., Schoendorf, A. and Croteau, R.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1999) Institute of Biological Chemistry, PO Box 646340, Pullman, WA 99164-6340, USA

FEATURES
source
1..1419
/organism="Taxus cuspidata"
/db_xref="taxon:99806"
1..1419
/gene="TAT"
11..1330
/note="TAT"
/note="transacylase"
/codon_start=1
/product="taxadienol acetyl transferase"
/protein_id="AAF34254.1"
/db_xref="GI:6978038"
/translation="MEKTDLHNLEKVGSPPLPKRTDLSIDNIPVRSIFN
ALLITNASPTMISADPAKPIREALAKLIVYPPAGLRTENGDEVECTGEGAM
ELEAMADNELSVLGFDDSPSFQQLSLPDLNFKDLISLVQVTRTCGFFVGV
SFHGVCDGRGAQFLKGLAEMARGEVKLSLPIVNNRELKLDDEPKYLOFHFEEFLRA
PSIVKEIVOTYFIIDFETINYIKOSMECKECSFEVSAAMTIAIRARQIPESE
YKILFGMDMRNSFNPPLSPGYSIGACVADVODLISGLIRATIMTIKKSIVSL
NDNEKRAVVKSELDVNNHNNVAFADMSLGDDEVDPCGNVSVSPVQOQSALA
MONTFLFKPSKPKPDGILIMFLPLSKKSKRIMEAMKRYVAKV"

BASE COUNT 430 a 262 c 311 g 416 t

ORIGIN

Query Match 83.5%; Score 14.2; DB 7; Length 1419;
Best Local Similarity 76.5%; Pred. No. 7.6e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAYTTTGNTGGGNA 17
11:11:11 111111
Db 1127 GATTTCGTTGGGNA 1143

RESULT 8
AB029340 1476 bp mRNA PLN 04-APR-2000
LOCUS Perilla frutescens mRNA for anthocyanin acyltransferase, partial
DEFINITION cds
AB029340
VERSION AB029340.1 GI:7415645
KEYWORDS anthocyanin acyltransferase.
SOURCE Perilla frutescens leaf cDNA to mRNA.
ORGANISM Perilla frutescens
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Lamiales; Lamiales; Perilla.

REFERENCE 1 (sites)
AUTHORS Sakakibara, K.Y., Tanaka, Y., Mizutani, M.F., Fujiwara, H., Fukui, Y., Ashikari, T., Yamaguchi, M. and Kusumi, T.
TITLE Molecular and biochemical characterization of a novel hydroxycinnamoyl-CoA: anthocyanin 3-O-glucoside-6
JOURNAL Unpublished (1999)
AUTHORS Sakakibara, K.Y. and Tanaka, Y.
TITLE Direct Submission
JOURNAL Submitted (24-JUN-1999) to the DDBJ/EMBL/GenBank databases. Keiko Y Sakakibara, Suntory Research Center, Fundamental Research, Plant

Biotech; Shimamoto-cho, Wakayamada, 1-1-1, Mishima-gun, Osaka 618-8503, Japan (E-mail: Keiko.Sakakibara@suntory.co.jp, Tel.: +81-75-962-8807, Fax: +81-75-962-8262)

FEATURES
source
1..1476
/organism="Perilla frutescens"
/db_xref="taxon:48386"
/lissue_type="leaf"
<1..1343
/codon_start=3
/product="anthocyanin acyltransferase"
/protein_id="BA93475.1"
/db_xref="GI:7415646"
/translation="VETCRVGPDPDSVAEQSVPLTFEDMTWLHFMLOLTFEPC
SKOHSESIVPKLOSSTKLHFPPLSCNLIYSPSPKMPERFRLSGDSVFTIIES
SDPDFDIVGNRESPVRLYVNPVPLPIYVESDRLFOVFAOVTLPGRGVGIAT
HHTVSDAPSEFLATWMSMSKHNEDDEDFKSLPVDRSVIKYIPKFDSTIYRNA
LKFPLQSRHPSLPTDRIKITTFVTSKIKKLGMQSRVPSLVHLSFVALATAMAG
ITKSTFADQDQDEDAFLIPVLRPLRPVPEVFGNCLVALPRMRRLVGEKG
VFLAAEVIAAEIKRINDKRILETVEKMSPEIRKALOKSYSPVAGSSKIDLGADFGW
GKARQETISIDGKRYAMTLCARPEGCLVCLSLPKMDAPAFAYFSLGICNG"

BASE COUNT 419 a 315 c 331 g 411 t

ORIGIN

Query Match 83.5%; Score 14.2; DB 7; Length 1476;
Best Local Similarity 76.5%; Pred. No. 7.6e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAYTTTGNTGGGNA 17
11:11:11 111111
Db 1167 GATTTCGATGGGNA 1183

RESULT 9
E12756 1479 bp DNA PAT 24-JUN-1998
LOCUS Perilla oclimoides mRNA for acyltransferase, partial cds.
DEFINITION E12756
ACCESSION E12756
VERSION E12756.1 GI:3251588
KEYWORDS JP 1997070290-A/4.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1479)
AUTHORS Ashikari, T., Tanaka, Y., Fujiwara, H., Nakao, M., Fukui, Y., Yonekura, K., Mizutani, M. and Kusumi, T.
TITLE GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY
JOURNAL Patent: JP 1997070290-A 4 18-MAR-1997;
SUNTORY LTD
OS Perilla oclimoides
PN JP 1997070290-A/4
PD 18-MAR-1997
PF 30-JAN-1996 JP 1996046534
PR 17-FEB-1995 JP 95P 67119, 29-JUN-1995 JP 95P 196915 PI
ASHIKARI TOSHIIKO, TANAKA YOSHIOKAZU, FUJIWARA HIROYUKI, PI NAKAO MASAHIRO,
PI FUKUI YUO, YONEKURA KEIKO, MIZUTANI MASAKO, KUSUMI TAKAKI PC
C12N15/09, A01H1/00, C07H21/04, C07K14/42, C12N9/10, (C12N9/10, PC
C12R1:865),
PC (C12N9/10, C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH key Location/Qualifiers
FH source 1..1479
FT /organism="Perilla oclimoides" FT
FT /lissue_type="leaves"
FT /clone="pSAT208"
FT CDS 1..1343
FT /product="acyltransferase".
FEATURES
source
1..1479
Location/Qualifiers

ORGANISM	REFERENCE
unidentified	1 (bases 1 to 1526)
unclassified.	

AUTHORS Aggelis,A., John,I., Karvouni,Z. and Grejerson,D.
TITLE FRUIT RIPENING
JOURNAL Patent: WO 9737023-A 09-OCT-1997;
ZENECA LTD (GB)
COMMENT Other publication AU 2167497 19971022.
FEATURES Location/Qualifiers
source 1. .1526
/organism="unidentified"
/db_xref="taxon:32644"
/clone="MEL 2"

BASE COUNT 439 a 297 c 331 g 450 t 9 others
ORIGIN

Query Match 83.5%; Score 14.2; DB 5; Length 1526;
Best Local Similarity 76.5%; Pred. No. 7.6e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGNTGGGNA 17
11:11:11 1111 11
Db 1136 GACTTGATGGGANA 1152

RESULT 13
116758 1542 bp DNA PAT 03-APR-1996
LOCUS Sequence 1 from patent US 5477002.
DEFINITION 116758
ACCESSION 116758
VERSION 116758.1 GI:1251666
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1542)
AUTHORS Tuttle,A.B. and Crossland,L.D.
TITLE Anther-specific CDNA sequences, genomic DNA sequences and
recombinant DNA sequences
JOURNAL Patent: US 5477002-A 1 19-DEC-1995;
FEATURES Location/Qualifiers
source 1. .1542
/organism="unknown"

BASE COUNT 473 a 295 c 330 g 444 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 5; Length 1542;
Best Local Similarity 76.5%; Pred. No. 7.6e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGNTGGGNA 17
11:11:11 1111 11
Db 1239 GATTTGGTGGGANA 1255

RESULT 14
NTHSR201 1563 bp mRNA PLN 12-SEP-1996
LOCUS N.tabacum mRNA for HSR201 protein.
DEFINITION X95343
ACCESSION X95343
VERSION X95343.1 GI:1171576
KEYWORDS hsr201 gene; hypersensitivity-related gene.
SOURCE Common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asterales; Solanales; Solanaceae; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 1563)
AUTHORS Czerwik,P., Huang,H.C. and Marco,Y.
TITLE Characterization of hsr201 and hsr515, two tobacco genes
preferentially expressed during the hypersensitive reaction
provoked by phytopathogenic bacteria
JOURNAL Plant Mol. Biol. 31 (2), 255-265 (1996)
MEDLINE 96343929

REFERENCE 2 (bases 1 to 1563)
AUTHORS Yves,M.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1996) M. Yves, UMR 05, CNRS-INRA, Laboratoire de
Biologie Molculaire, Chemin de Borde Rouge, BP27, 31326
Castanet-Tolosan Cedex, France
FEATURES Location/Qualifiers
source 1. .1563
/organism="Nicotiana tabacum"
/cultivar="bottom spectral"
/db_xref="taxon:4097"
/dev_stage="adult"
/clone_1lb="PBK phagemid vector"
/tissue_type="leaf"
1. .1563
/evidence="experimental"
40. .1422
/gene="hsr201"
40. .1422
/gene="hsr201"
/note="hypersensitivity-related gene"
/codon_start=1
/protein_id="CA64636.1"
/db_xref="GI:1171577"
/db_xref="SPTREMBL:O43583"
/translation="MDSKOSSELPVTVRORPELIAPAKPTPRETKFLSDIDDEGLR
FOIPVIOFYKDSMSGRDPYVIRKIAETLVFYYPAGRLREGNRLMWDGEG
IMFVADADVTLEORGDLEPPCLLELIDVDPSCAGVILNGLITVTRLCGGFI
PALRLNTHMSDAPGLVQFMTAVANGENARGSGSILPYNCRELILNRPVYTCRTHY
DEVROTGTITPLDDVHKSFEFGFSEVSALEKRVPHLRKSTELTJAVLMCRITM
SLKPDPEEVRALCIIVNARSFNPLPYGTGNAAPVAATYTAARLSKNPLGVALEL
VKTKSDVTEEYEMKSVADLMVLRGPHFTVVTGTVFVSDVTRGCEGEVPGMGKAYVGG
PAKGGVGAIPGVASFYIPFKNKGKGIIVPICLPGFAMETPVKELDMLKVDAPLVN
SNVATIRPAL"

polya_site 1563
/evidence="experimental"

BASE COUNT 430 a 305 c 344 g 484 t
ORIGIN

Query Match 83.5%; Score 14.2; DB 48; Length 1563;
Best Local Similarity 76.5%; Pred. No. 7.6e+02;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAYTTYGNTGGGNA 17
11:11:11 1111 11
Db 1183 GATTTGGATGGGANA 1199

RESULT 15
E12755 1605 bp DNA PAT 24-JUN-1998
LOCUS Petunia hybrida mRNA for acyltransferase,complete cds.
DEFINITION E12755
ACCESSION E12755
VERSION E12755.1 GI:3251587
KEYWORDS JP 1997070290-A/3.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1605)
AUTHORS Ashikari,T., Tanaka,Y., Fujiiwara,H., Nakao,M., Fukui,Y.,
Yonekura,K., Mizutani,M. and Kusumi,T.
TITLE GENE CODING PROTEIN HAVING ACYL GROUP TRANSFER ACTIVITY
JOURNAL Patent: JP 1997070290-A 3 18-MAR-1997;
COMMENT OS Petunia hybrida
PN JP 1997070290-A/3
PD 18-MAR-1997
PR 30-JAN-1996 JP 1996046534
PF 17-FEB-1995 JP 95P 67159, 29-JUN-1995 JP 95P 196915 PI
ASHIKARI TOSHIHIKO, TANAKA YOSHIKAZU, FUJIIWARA HIROYUKI, PI NAKAO
MASAHIRO,
PI FUKUI YUKO, YONEKURA KEIRO, MIZUTANI MASAKO, KUSUMI TAKANAKI PC

C12N15/09, A01H1/00, C07H21/04, C07K14/42, C12N9/10, (C12N9/10, PC

C12R1:865),

PC (C12N9/10, C12R1:19);

CC strandedness: Double;

CC topology: Linear;

FT Key Location/Qualifiers

FT source

1. .1605

FT /organism='Petunia hybrida'

FT /tissue_type='petal'

FT /clone='pPAT48'

FT CDS 67. .1413

FT /product='acyltransferase'

FT Location/Qualifiers

1. .1605

/organism='unidentified'

/db_xref='taxon:32644'

BASE COUNT 477 a 318 c 376 g 434 t

ORIGIN

FEATURES

Source

1. .1605

/organism='unidentified'

/db_xref='taxon:32644'

BASE COUNT 477 a 318 c 376 g 434 t

ORIGIN

FEATURES

Source

1. .1605

/organism='unidentified'

/db_xref='taxon:32644'

BASE COUNT 477 a 318 c 376 g 434 t

ORIGIN

FEATURES

Source

1. .1605

/organism='unidentified'

/db_xref='taxon:32644'

BASE COUNT 477 a 318 c 376 g 434 t

ORIGIN

FEATURES

Source

1. .1605

/organism='unidentified'

/db_xref='taxon:32644'

BASE COUNT 477 a 318 c 376 g 434 t

ORIGIN

FEATURES

Source

1. .1605

/organism='unidentified'

/db_xref='taxon:32644'

BASE COUNT 477 a 318 c 376 g 434 t

ORIGIN

FEATURES

Source

1. .1605

/organism='unidentified'

Search completed: October 20, 2000, 03:37:59

Job time: 7317 sec

Query Match 83.5%; Score 14.2; DE 5; Length 1605;

Best Local Similarity 76.5%; Pred. No. 7.5e+02;

Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGCGTGGCGNAA 17

Db 1234 GATTTGCGTGGCGNAA 1250

THIS PAGE BLANK (USPTO)

ADDRESSEE: Flinnegan, Henderson, Farbow, Garrett &
ADDRESSSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,349
FILING DATE: 25-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/364,439
FILING DATE:
APPLICATION NUMBER: US 08/112,061
FILING DATE: 26-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barker, M. P.
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 02481.1323-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3581 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Mus musculus
STRAIN: osteoblastic cell line MC3T3E1
FEATURE:
NAME/KEY: CDS
LOCATION: 284..2671
US-08-738-349-1

```

Query Match          2.3%:  Score 39.8; DB 3.  Length 3581;
Best Local Similarity 79.7%; Pred. No. 0.12;
Matches 47; Conservative 0; Mismatches 12; Indels 0; Gaps 0.

Oy  1645 AAGTATCCTATGCAATAGACAACATGTTATGTGTTAAAAAAAAAAAAAAAAAAAA 1703
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   3469 AAGCAACCTTGAATAATMAAAAAAAAAAGATCTCTTTTAAAAAAAAAAAAAAAAAAAA 3527

RESULT 3
US-08-021-608D-9
Sequence 9, Application US/08021608D
Patent No. 5580760
GENERAL INFORMATION:
APPLICANT: LEVENS, DAVID L., DUNCAN,
APPLICANT: ROBERT C., AND AYIGAN, MARK I.
TITLE OF INVENTION: NOVEL FUSE BINDING
TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE

```

```

1      OPERATING SYSTEM: PC-DOS/MS-DOS
2      SOFTWARE: WORDPERFECT 5.1
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER: US/08/021,608D
5      FILING DATE: 22-FEB-1993
6      CLASSIFICATION: 435
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER:
9      FILING DATE:
10     CLASSIFICATION: 435
11     ATTORNEY/AGENT INFORMATION:
12     NAME: WILLIAM S. FEILER
13     REGISTRATION NUMBER: 26,728
14     REFERENCE/DOCKET NUMBER: 2026-4063
15     TELECOMMUNICATION INFORMATION:
16     TELEPHONE: (212) 751-8400
17     TELEFAX: (212) 751-6849
18     INFORMATION FOR SEQ ID NO: 9:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 2381
21     TYPE: Nucleic acid
22     STRANDEDNESS: Double
23     TOPOLOGY: Unknown
24     MOLECULE TYPE: cDNA
25     HYPOTHEICAL: NO
26     ANTI-SENSE: YES
27     ORIGINAL SOURCE:
28     ORGANISM: Human
29     CELL LINE: HL60
30     FEATURE:
31     OTHER INFORMATION: 470 bp variable
32     OTHER INFORMATION: region where R is A or G
33
34 US-08-021-6080-9

```

```

Query Match Similarity      2.3%:  Score 38.4;  DB 1;  Length 2381;
Best Local Similarity      64.8%:  Pred. No. 0.24;
Matches      57;  Conservative      0;  Mismatches      31;  Indels      0;  Gaps      0.

Cy      1616  ATTGTAATAATGCCATTATATACTTCGCATTAAGATATCTATGCAATAGACATGTTATG 1675
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2260  AATGTAATACCTTTCATCTTCCTCAATGCGTGTTTGTCCTTACATAAATGATATGAAACC 2319

Cy      1676  TGTATAAAAAAAAAAAAAAAAAAAAAA 1703
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2320  TCATAAAAAAAAAAAAAAAAAAAAAA 2347

RESULT      4
US-08-726-160-9
: Sequence 9, Application US/08726160
: Patent No. 5734016
: GENERAL INFORMATION:
: APPLICANT:  LEVENS, DAVID L., DUNCAN,
: APPLICANT:  ROBERT C., AND AVIGAN, MARK I.
: TITLE OF INVENTION:  NOVEL FUSE BINDING
: TITLE OF INVENTION:  PROTEIN AND CDNA THEREFOR
: NUMBER OF SEQUENCES:  24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE:  MORGAN & FINNEGAN
: STREET:     345 PARK AVENUE
: CITY:       NEW YORK
: STATE:     NEW YORK
: COUNTRY:   USA
: ZIP:       10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE:  FLOPPY DISK
: COMPUTER:     IBM PC COMPATIBLE
: OPERATING SYSTEM:  PC-DOS/MS-DOS
: SOFTWARE:     WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER:  US/08/726.160
: FILING DATE:      04-OCT-1996

```



```

NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: YES
ORIGINAL SOURCE:
ORGANISM: Human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE: HL60
ORGANELLE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: 470 bp variable
PCT-US94-01782-9
region where R is A or G.

Query Match      2.3%; Score 38.4; DB 6; Length 2381;
Best Local Similarity 64.8%; Pred.No. 0.24;
Matches 57; Conservative 0; Mismatches 31; Indels 0; Gaps 0

QY 1616 ATTGATTAATGCCATTATATCTTCCATAAAGTATCCTATGCATATGACAAATGTATTG 1675
Db 2260 AATGATATCTTTCACTTCCTCCAATGCCCTGTTTTGGCTTTACAATAATGAATATGAACC 2319

QY 1676 TGTTAACAAAAAAAAAAAAAAAAAAAAAAA 1703
Db 2320 TCACAAAAAAAAAAAAAAAAAAAAAAA 2347

RESULT 6
US-08-021-608D-1
Sequence 1, Application us/08021608D
Patent No. 5580760
GENERAL INFORMATION:
APPLICANT: LEVENS, DAVID L., DUNCAN,
APPLICANT: ROBERT C., AND AVIGAN, MARK I.
TITLE OF INVENTION: NOVEL FUSE BINDING
TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FILOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/021.608D
: FILING DATE: 22-FEB-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: WILLIAM S. FEILER
: REGISTRATION NUMBER: 26,728
: REFERENCE/DOCKET NUMBER: 2026-4063
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 758-4800
: TELEFAX: (212) 751-6849
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2384
: TYPE: Nucleic acid
: STRANDEDNESS: Double
: TOPOLOGY: Unknown
: MOLECULE TYPE: CDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: YES
: ORIGINAL SOURCE:
: ORGANISM: Human
: CELL LINE: HL60
: FEATURE:
: OTHER INFORMATION: 473 bp
: OTHER INFORMATION: variable region where R is A or G.
: US-08-021-608D-1

Query Match      2.3%; Score 38.4; DB 1; Length 2384;
Best Local Similarity 64.8%; Pred. No. 0.24; Mismatches 0; Gaps 0;
Matches 57; Conservative 0; Indels 31;

Oy 1616 ATTGATATAGCCATTATATCTTCATTAAGTATCCTATGCAATAGAGAACATGTATG 1675
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2263 AATGATATCTTCACTTCCAAATGCCGTGTTGCTTACAAATATGATATGAAC 2322
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy 1676 TGTAAAAAAAAAAAAAAAAAAAAA 1703
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2323 TCAAAAAAAAAAAAAAAAAAAAAA 2350
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-08-726-160-1
: Sequence 1, Application US/08726160
: Patent No. 5734016
: GENERAL INFORMATION:
: APPLICANT: LEVENS, DAVID L., DUNCAN,
: APPLICANT: ROBERT C., AND AVIGAN, MARK I.
: TITLE OF INVENTION: NOVEL FUSE BINDING
: TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/726,160
: FILING DATE: 04-OCT-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
```

```

: APPLICATION NUMBER: 08/021,608
: FILING DATE: 22-FEB-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: WILLIAM S. FEILER
: REGISTRATION NUMBER: 26,728
: REFERENCE/DOCKET NUMBER: 2026-4063U51
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 758-4800
: TELEFAX: (212) 751-6849
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2384
: TYPE: Nucleic acid
: STRANDEDNESS: Double
: TOPOLOGY: Unknown
: MOLECULE TYPE: CDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: YES
: ORIGINAL SOURCE:
: ORGANISM: Human
: CELL LINE: HL60
: FEATURE:
: OTHER INFORMATION: 473 bp
: OTHER INFORMATION: variable region where R is A or G.
: US-08-726-160-1

Query Match      2.3%; Score 38.4; DB 2; Length 2384;
Best Local Similarity 64.8%; Pred. No. 0.24; Mismatches 0; Gaps 0;
Matches 57; Conservative 0; Indels 31;

Oy 1616 ATTGATATAGCCATTATATCTTCATTAAGTATCCTATGCAATAGAGAACATGTATG 1675
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2263 AATGATATCTTCACTTCCAAATGCCGTGTTGCTTACAAATATGATATGAAC 2322
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy 1676 TGTAAAAAAAAAAAAAAAAAAAAA 1703
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2323 TCAAAAAAAAAAAAAAAAAAAAAA 2350
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
PCT-US94-01782-1
: Sequence 1, Application PC/TUS9401782
: GENERAL INFORMATION:
: APPLICANT: THE GOVERNMENT OF THE UNITED STATES
: APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
: APPLICANT: HEALTH AND HUMAN SERVICES
: TITLE OF INVENTION: NOVEL FUSE BINDING
: TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/01782
: FILING DATE: 22-FEB-1994
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. 08/021,608
: FILING DATE: 22-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: WILLIAM S. FEILER
: REGISTRATION NUMBER: 26,728
```

```

? APPLICATION NUMBER: US/08/684,862
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/361,705
? FILING DATE:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/966,040
? FILING DATE: 30-DEC-1992
? APPLICATION NUMBER: PCT/EP91/01361
? FILING DATE: 19-JUL-1991
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 988 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA to mRNA
? ORIGINAL SOURCE:
? ORGANISM: Agkistrodon rhodostoma
? FEATURE:
? LOCATION: 197 to 904
? OTHER INFORMATION: the coding region shown in (2){(x)}(B)
? OTHER INFORMATION: codes for the protein of SEQ ID NO: 5
US-08-684-862-10

Query Match                2.2%   Score 37.2; DB 2; Length 988;
Best Local Similarity      77.6%; Pred. No. 0.35;
Matches    45; Conservative    0; Mismatches    13; Indels    0; Gaps    0.

QY  1646 AGTATCCTATGCATAGACACATGTTATGTATAAAAAAAAAAAAAAAAAAAAAA 1703
     ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   931 AGTTCCAAAGAAGATTAAACTTAATTAATGCTGTTAAAAA
                                     AAAAAAAAAAAAAAAAAAAAAA 988

RESULT 10
US-09-065-474-102
? Sequence 102, Application US/09065474
? Patent No. 6063599
? GENERAL INFORMATION:
? APPLICANT: Tang, E. Liang
? TITLE OF INVENTION: DIOFILARIA AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? TITLE OF INVENTION: USES THEREOF
? NUMBER OF SEQUENCES: 171
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Carol Talkington Verser, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: Windows 95
? SOFTWARE: Wordperfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/065,474
? FILING DATE: 24-APR-1998
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 37,459
? REFERENCE/DOCKET NUMBER: HW-5-C1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 102:
? SEQUENCE CHARACTERISTICS:
```

LENGTH: 184 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..183
US-09-065-474-102

Query Match 2.1%, Score 36.6; DB 5; Length 184;
Best Local Similarity 57.4%; Pred. No. 0.25;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1042 CACATGAAGAGTAAAGGGGATTAAGGGCTTCTTGTCAGAGTGCACCTATTGGAGAAG 1101
DB 26 CACAAGAACCGTTTTCAGCAGTAGGATTAATGGACACCTGAGATCCAGAAGAAC 85
QY 1102 CCATTGAAAGAGGTTGCAACAAGAAAGGCGTTCTTGACAGATGCAAAAACCTTG 1156
DB 86 CCAAGAAGACGATCATTCACGACGAAGAGAGAGAGTTCGACTTTCAGAAATTGCG 140

RESULT 11
US-09-065-474-103/C
Sequence 103, Application US/09065474
Patent No. 6063599

GENERAL INFORMATION:

APPLICANT: Tang, Liang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
NUMBER OF SEQUENCES: 171
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/065,474
FILING DATE: 24-APR-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 103:

SEQUENCE CHARACTERISTICS:
LENGTH: 184 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-065-474-103

Query Match 2.1%, Score 36.6; DB 5; Length 184;
Best Local Similarity 57.4%; Pred. No. 0.25;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1042 CACATGAAGAGTAAAGGGGATTAAGGGCTTCTTGTCAGAGTGCACCTATTGGAGAAG 1101
DB 159 CACAAGAACCGTTTTCAGCAGTAGGATTAATGGACACCTGAGATCCAGAAGAAC 100
QY 1102 CCATTGAAAGAGGTTGCAACAAGAAAGGCGTTCTTGACAGATGCAAAAACCTTG 1156
DB 99 CCAAGAAGACGATCATTCACGACGAAGAGAGAGAGTTCGACTTTCAGAAATTGCG 45

RESULT 12
US-09-031-485-9

Sequence 9, Application US/09031485
Patent No. 5824306
GENERAL INFORMATION:

APPLICANT: Tang, Liang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,485
FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/847,429
FILING DATE: 24-APR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 810 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-031-485-9

Query Match 2.1%, Score 36.6; DB 2; Length 810;
Best Local Similarity 57.4%; Pred. No. 0.47;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1042 CACATGAAGAGTAAAGGGGATTAAGGGCTTCTTGTCAGAGTGCACCTATTGGAGAAG 1101
DB 224 CACAAGAACCGTTTTCAGCAGTAGGATTAATGGACACCTGAGATCCAGAAGAAC 283
QY 1102 CCATTGAAAGAGGTTGCAACAAGAAAGGCGTTCTTGACAGATGCAAAAACCTTG 1156
DB 284 CCAAGAAGACGATCATTCACGACGAAGAGAGAGAGTTCGACTTTCAGAAATTGCG 338

RESULT 13

US-09-031-485-10/C
Sequence 10, Application US/09031485

Patent No. 5824306
GENERAL INFORMATION:
APPLICANT: Tang, Llang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Port Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,485
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/847,429
FILING DATE: 24-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-031-485-10

Query Match 2.1%; Score 36.6; DB 2; Length 810;
Best Local Similarity 57.4%; Pred. No. 0.47;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1042 CACATAAAGAGTAGTTGGGATAAAGGCGTCTTGTTCGAGTTGCAGCTATTGGAGAG 1101
DB 587 CACAAAGACCGTTTTCACAGTAGGCTATATAGGACACCTGGAGATCCAGAAAGAC 528

QY 1102 CCATTGAAAAGAGTTCACACAAAGAAAGCGCTTCCTGCGAGATGCAAAAACCTTG 1156
DB 527 CCAAGAACAGTCATTCACGAAGAGAGAGAGAGTTCGAGATTTCAGAAATTGCG 473

RESULT 14
US-08-847-429A-9
Sequence 9, Application US/08847429A
Patent No. 5827692
GENERAL INFORMATION:
APPLICANT: Tang, Llang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive

CITY: Port Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/847,429A
FILING DATE: 24-APR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-847-429A-9

Query Match 2.1%; Score 36.6; DB 2; Length 810;
Best Local Similarity 57.4%; Pred. No. 0.47;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1042 CACATAAAGAGTAGTTGGGATAAAGGCGTCTTGTTCGAGTTGCAGCTATTGGAGAG 1101
DB 224 CACAAAGACCGTTTTCACAGTAGGCTATATAGGACACCTGGAGATCCAGAAAGAC 283

QY 1102 CCATTGAAAAGAGTTCACACAAAGAAAGCGCTTCCTGCGAGATGCAAAAACCTTG 1156
DB 284 CCAAGAACAGTCATTCACGAAGAGAGAGAGAGTTCGAGATTTCAGAAATTGCG 338

RESULT 15
US-08-847-429A-10/C
Sequence 10, Application US/08847429A
Patent No. 5827692
GENERAL INFORMATION:
APPLICANT: Tang, Llang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Port Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/847,429A
FILING DATE: 24-APR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington

